MASTER EN SCIENCES MATHÉMATIQUES

Méthodes de points intérieurs du type gradient et -sous-gradient pour la minimisation avec contrainte convexe

Bonduel, Lieve

Award date:
2007

FUNDP
NAMUR

Faculté des Sciences
Département de Mathématique

Rempart de la Vierge, 8
B - 5000 Namur (Belgique)

Méthodes de points intérieurs du type gradient et ϵ-sous-gradient pour la minimisation avec contrainte convexe.

Mémoire présenté pour l'obtention
du grade de
Licencié en Sciences Mathématiques
par

BONDUEL **Lieve**

Promoteur : STRODIOT Jean-Jacques

Année Académique 2006-2007

Résumé.

Méthodes de points intérieurs du type gradient et ϵ-sous-gradient pour la minimisation avec contrainte convexe.

Nous étendons les méthodes du $\epsilon-$sous-gradient pour les problèmes de minimisation non différentiables sans contrainte aux problèmes avec contrainte sur un ensemble polyédral, en particulier sur $\mathbb{R}_+^p$. Nous faisons cela en remplaçant le terme de régularisation quadratique dans le schéma du sous-gradient par la fonction distance logarithmique quadratique. Nous obtenons ainsi les méthodes intérieures du $\epsilon-$sous-gradient, qui étendent naturellement la méthode de faisceaux et de la méthode de projection du sous-gradient de Polyak pour la minimisation convexe non différentiable. De plus, des extensions similaires sont considérées pour la minimisation avec contrainte différentiable pour fournir des méthodes de descente du gradient intérieur.

Mots clés : optimisation convexe non différentiable, algorithmes du sous gradient, méthode faisceaux, méthode proximale logaritmique quadratique, méthode du gradient projeté

Interior gradient and $\epsilon-$subgradient descent methods for constrained convex minimisation.

We extend epsilon-subgradient descent methods for unconstrained nonsmooth convex minimization to constrained problems over polyhedral sets, in particular over $\mathbb{R}_+^p$. This is done by replacing the usual squared quadratic regularization term used in subgradient schemes by the logarithmic-quadratic distancelike function. We then obtain interior $\epsilon-$subgradient descent methods, which allow us to provide a natural extension of bundle methods and Polyak's subgradient projection methods for nonsmooth convex minimization. Furthermore, similar extensions are considered for smooth constrained minimization to produce interior gradient descent methods.

Keys words : nondifferentiable convex optimization, subgradient algorithms, bundle methods, logarithmic-quadratic proximal methods, projected gradient methods

Remerciements.

Je tiens à remercier Monsieur Strodiot, promoteur de mon travail, qui s'est toujours rendu disponible pour me renseigner, me conseiller et répondre à mes questions.

Je tiens également à remercier mes parents et ma famille de m'avoir soutenu tout au long de mes études.

Table des matières

Introduction.

Dans ce mémoire, nous considérons le problème de minimisation convexe suivant :

$$(P) \qquad f_\star = \inf\{f(x) : x \in C\},$$

où $f : \mathbb{R}^p \to \mathbb{R}\cup\{+\infty\}$ est une fonction propre, semi-continue-inférieurement et convexe et où C est un ensemble polyédral, et en particulier l'orthant positif.

Les méthodes classiques pour résoudre (P) quand f n'est pas nécessairement différentiable (ou deux fois différentiables) sont les méthodes du gradient (sous-gradient) projeté et de faisceaux. Pour ces deux méthodes, la contrainte $x \in \mathbb{R}^p_+$ mène a des problèmes numériques considérables.

Essayons d'en expliquer les raisons. Premièrement, considérons le cas sans contrainte. Alors, le schéma de la méthode gradient /sous-gradient est donné par :

$$x^0 \in \mathbb{R}^p, \quad x^k = x^{k-1} - \lambda_k d^k, \text{ où } d^k = \nabla f(x^{k-1}) \text{ ou } d^k = c^k \in \partial f(x^{k-1}),$$

où $\partial f(x^{k-1})(\nabla f(x^{k-1}))$ est le sous-différentiel (gradient) de f en x^{k-1}. Un tel schéma est raisonnable. En effet, d^k est une direction de descente naturelle pour f au point x^{k-1}, alors que c^k a les mêmes propriétés pour la fonction $x \to d(x, X_\star)$, la distance de x à l'ensemble optimal $X_\star$. Maintenant, considérons le cas avec contraintes, par exemple, $C = \mathbb{R}^p_+$. Alors les méthodes sont étendues en prenant comme point suivant la projection euclidienne de x^k sur $\mathbb{R}^p_+$. Malheureusement, la projection euclidienne détruit la propriété de descente et peut être à l'origine de zigzags, malgré la théorie de convergence. L'extension de la méthode faisceaux pour l'orthant positif nous conduit aussi à considérer un sous-problème qui consiste à minimiser une fonction quadratique sur l'intersection de $\mathbb{R}^p_+$ et d'un polyèdre, un problème clairement plus compliqué que le cas sans contrainte.

La question que l'on se pose alors est : peut-on construire un algorithme de type gradient ou sous-gradient pour le problèmes de minimisation avec contrainte (P) tel que les difficultés décrites ci-dessus soient éliminées ?

L'agencement de ce travail est le suivant : dans un premier chapitre, nous donnons quelques définitions et propriétés utiles pour la suite. Le second chapitre développe la méthode du point proximal. Dans le chapitre suivant nous parlons de la méthode proximale logarithmique-quadratique. Le chapitre quatre étendra la méthode de faisceaux à la méthode de faisceaux intérieure. Le cinquième chapitre traite les méthodes de descente du ϵ-sous gradient. Et le dernier chapitre nous parle de la méthode intérieure du sous-gradient.

Ce travail est basé essentiellement sur les articles [1] et [3] de Auslender et Teboulle.

Chapitre 1

Préliminaires.

1.1 Quelques définitions.

Définition 1.1.1 *Soit $f : I\!R^n \to I\!R \cup \{+\infty\}$. Le domaine de f est l'ensemble*

$$dom\ f = \{x \in I\!R^n | f(x) < +\infty\}.$$

La fonction f est propre si son domaine est non vide.

Définition 1.1.2 *Soit $f : I\!R^n \to I\!R \cup \{+\infty\}$ une fonction propre. L'épigraphe de f est l'ensemble non vide*

$$epi\ f = \{(x, r) \in I\!R^n \times I\!R \,|\, r \geq f(x)\}.$$

f est dite être convexe si son épigraphe est un sous-ensemble convexe de $I\!R^n \times I\!R$.

1.1.1 Les fonctions convexes fermées.

Définition 1.1.3 *Soit $f : I\!R^n \to I\!R \cup \{+\infty\}$ une fonction propre. On dit que f est semicontinue inférieurement (s.c.i) en $x \in I\!R^n$ si*

$$\lim \inf_{y \to x} f(y) \geq f(x) \quad [\Leftrightarrow \forall \{y_k\} \to x \lim \inf_{k \to \infty} f(y_k) \geq f(x)]$$

On a aussi que f est s.c.i si l'épigraphe de f est un ensemble fermé de $\mathbb{R}^n \times \mathbb{R}$.

Définition 1.1.4 *La fonction $f : I\!R^n \to I\!R \cup \{+\infty\}$ est fermée si son épigraphe est un sous-ensemble fermé de $I\!R^n \times I\!R$.*

1.1.2 Les fonctions affines.

Définition 1.1.5 *Soit $a \in I\!R^n$ et $b \in I\!R$. La fonction $f : I\!R^n \to I\!R$ définie par $f(x) = \langle a, x\rangle - b \quad \forall x \in I\!R^n$ est une fonction affine sur $I\!R^n$. Si $b = 0$, alors $f(x) = \langle a, x\rangle$ est une fonction linéaire.*

On a que les fonctions affines sont des fonctions convexes et fermées.

1.1.3 Les fonctions conjuguées.

Définition 1.1.6 *Soit $f : I\!R^n \to I\!R \cup \{+\infty\}$ une fonction propre minorée par une fonction affine. Alors la fonction $f^\star : I\!R^n \to I\!R \cup \{+\infty\}$ définie par*

$$f^\star(x) = \sup_{y \in I\!R^n} \{\langle x, y\rangle - f(y)\}$$

est la fonction conjuguée de f.

Proposition 1.1.1 *Soit $f : I\!R^n \to I\!R \cup \{+\infty\}$ une fonction propre minorée par une fonction affine. Alors la conjuguée de la fonction f est bien définie, propre, fermée et convexe.*

Preuve :

La fonction $f^\star$ est bien définie parce que f est propre. De plus, son domaine est non-vide. En effet, puisque f est minorée par une fonction affine, il existe $x_0 \in \mathbb{R}^n$ et $r \in \mathbb{R}$ tel que $f(y) \geq \langle x_0, y\rangle - r$ pour tout $y \in \mathbb{R}^n$. Donc, $f^\star(x_0) \leq r$ et le domaine de $f^\star$ est non vide. Finalement, puisque chaque fonction $x \mapsto \langle x, y\rangle - f(y)$ est affine, il suit que $f^\star$ est une fonction convexe propre et fermée. $\square$

1.1.4 La fonction de récession.

Définition 1.1.7 *Soit f une fonction propre, convexe. On définit f_∞ la fonction de récession de f par*

$$\forall y \quad (f_\infty)(y) = \sup\{f(x+y) - f(x) | \, x \in dom\ f\}$$

Propriétés :

- Si f est fermée alors f_∞ est fermée.
- On a une autre caractérisation de f_∞

$$\forall x \in \text{dom } f \quad (f_\infty)(y) = \sup_{\lambda > 0} \frac{f(x+\lambda y) - f(x)}{\lambda} = \lim_{\lambda \to \infty} \frac{f(x+\lambda y) - f(x)}{\lambda}$$

1.1.5 Le dual de Fenchel.

Le Théorème de dualité de Fenchel permet de résoudre des problèmes de minimisation d'une fonction du type $f(x) - g(x)$ où f est une fonction propre convexe sur $\mathbb{R}^n$ et g est une fonction propre concave sur $\mathbb{R}^n$.

Définition 1.1.8 *Soit $f : X \to I\!R \cup \{+\infty\}$, $g : Y \to I\!R \cup \{+\infty\}$, $A : X \to Y$ linéaire, continue.*
Considérons le problème

$$(v) \quad \min_{x \in X} f(x) + g(Ax)$$

Alors le dual de Fenchel est défini comme

$$(v^\star) \quad \min_{q \in Y^\star} f^\star(-A^\star q) + g^\star(q)$$

Voici maintenant le Théorème de dualité de Fenchel.

Théorème 1.1.1 *Soit f une fonction propre, convexe de $I\!R^n$, et soit g une fonction propre, concave de $I\!R^n$. On a aussi*

$$\inf_x \{f(x) - g(x)\} = \sup_{x^\star} \{g^\star(x^\star) - f^\star(x^\star)\}$$

si l'une des conditions suivantes suit :

(a) int (dom f) $\cap$ int (dom g) $\neq \emptyset$;

(b) f et g sont fermées et int (dom $f^\star$) $\cap$ int (dom $g^\star$) $\neq \emptyset$.

Sous (a) le supremum est atteint en un $x^\star$,et sous (b) l'infinimum est atteint en un x; si nous avons (a) et (b) alors, l'infinimum et le suprémum sont finis.

Preuve :

Voir [8] Théorème 31.1 p 327 □

1.2 Le sous-différentiel.

Une fonction f peut être convexe sans être différentiable (par exemple : $f(x) = |x|$ qui est non différentiable en 0). C'est pour cela que nous définissons les ensembles sous-différentiels.

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe et $x_0 \in \mathbb{R}^n$. Nous savons que la dérivée directionnelle $f'(x_0, \cdot)$ est une fonction sous-linéaire et à valeur finie. Donc $f'(x_0, \cdot)$ est le support de l'ensemble compact, convexe, non vide

$$S = \{s \in \mathbb{R}^n | \langle s, d \rangle \leq f'(x_0, d) \, \forall d \in \mathbb{R}^n\}.$$

Définition 1.2.1 *Le sous-différentiel $\partial f(x)$ de f en x est l'ensemble compact convexe non vide de $I\!R^n$ dont le support est la fonction $f'(x, \cdot)$, c'est-à-dire*

$$\partial f(x) = \{s \in I\!R^n | \langle s, d\rangle \ \leq f'(x, d) \quad \forall d \in I\!R^n\}.$$

N'importe quel vecteur $s \in \partial f(x)$ est appelé sous-gradient de f en x.

Le sous-différentiel peut aussi être caractérisé par une inégalité.

Proposition 1.2.1 *Nous avons la propriété suivante :*

$$s \in \partial f(x_0) \Leftrightarrow \forall x \in I\!R^n \quad f(x) \geq f(x_0) + \langle s, x - x_0\rangle.$$

Preuve :

$$\begin{aligned}
s \in \partial f(x_0) &\Leftrightarrow \langle s, d\rangle \leq f'(x, d) \quad \forall d \in \mathbb{R}^n \\
&\Leftrightarrow \langle s,\, d\rangle \leq \inf_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} \quad \forall d \in \mathbb{R}^n \\
&\Leftrightarrow \langle s,\, d\rangle \leq \frac{f(x_0 + td) - f(x_0)}{t} \quad \forall t > 0, \forall d \in \mathbb{R}^n \\
&\Leftrightarrow f(x_0 + td) \geq f(x_0) + \langle s,\, td\rangle \quad \forall t > 0, \forall d \in \mathbb{R}^n \\
&\Leftrightarrow f(x) \geq f(x_0) + \langle s, x - x_0\rangle \quad \forall x \in \mathbb{R}^n.
\end{aligned}$$

□

Proposition 1.2.2 *Soit $f : I\!R^n \to I\!R$ une fonction convexe et soit $x_0 \in I\!R^n$. Nous avons alors l'équivalence entre les deux conditions suivantes.*

(i) x_0 *minimise* f *sur* $I\!R^n$;

(ii) $0 \in \partial f(x_0)$.

Preuve :

Par la définition du sous-differentiel, il est immédiat que (i) est équivalent à (ii). □

Proposition 1.2.3 *La multi-fonction ∂f est localement bornée dans le sens où pour chaque sous ensemble borné B de $I\!R^n$,*

$$\partial f(B) = \cup_{b \in B} \partial f(b) \text{ est borné dans } I\!R^n.$$

Preuve :

Voir [8] Théorème 24.7 $\square$

1.3 L'approximation du sous-différentiel.

Définition 1.3.1 *Soit f propre, convexe, s.c.i, $\epsilon \geq 0$ et $x \in dom\, f$. Un vecteur $s \in I\!R^n$ est appelé un ϵ-sous-gradient de f en x si*

$$\forall y \in dom\, f \quad f(y) \geq f(x) + \langle s, y - x\rangle - \epsilon.$$

L'ensemble des ϵ-sous-gradients de f en x est noté par $\partial_\epsilon f(x)$ et appelé le ϵ-sous-différentiel de f en x.
Par convention, $\partial_\epsilon f(x) = \emptyset$ quand $x \notin dom\, f$.

Proposition 1.3.1 *Soit $\epsilon > 0$. Alors $\partial_\epsilon f(x) \neq \emptyset$ pour tout $x \in dom\, f$.*

Preuve :

Voir [6] Théorème 1.1.2 $\square$

Proposition 1.3.2 *Formule de transport*
Soit $x, y \in dom\, f$ et soit $s(y) \in \partial f(y)$. Alors

$$s(y) \in \partial_\epsilon f(x) \Leftrightarrow f(y) \geq f(x) + \langle s(y), y - x\rangle - \epsilon.$$

Preuve :

$\Rightarrow$: évident.
$\Leftarrow$: Soit $x, y, z \in$ dom f. Alors, utilisant successivement la définition du sous-gradient et l'hypothèse, nous obtenons

$$\begin{aligned} s(y) \in \partial f(y) &\Rightarrow f(z) \geq f(y) + \langle s(y),\, z - y\rangle \\ &\Rightarrow f(z) \geq f(x) + \langle s(y),\, y - x\rangle - \epsilon + \langle s(y),\, z - y\rangle \\ &\Rightarrow f(z) \geq f(x) + \langle s(y),\, z - x\rangle - \epsilon. \end{aligned}$$

D'où $s(y) \in \partial_\epsilon f(x)$ et la preuve est complète. $\square$

Chapitre 2

La méthode du point proximal.

2.1 La régularisée de Moreau-Yosida.

Définition 2.1.1 *Soit f propre, convexe, s.c.i et $c > 0$. La fonction F : $I\!R^n \to I\!R$ définie par*

$$F(x) = \min_{y \in I\!R^n} \{f(y) + \frac{c}{2}\|y - x\|^2\}$$

est appelée la régularisée de Moreau-Yosida de f.

Ici, nous considérons une fonction f qui peut prendre des valeurs égales à $+\infty$. Ceci nous permet d'examiner les problèmes de la forme :

$$\min\{h(x) | \, x \in C\}$$

où h : $\mathbb{R}^n \to \mathbb{R}$ une fonction convexe et $C \subseteq \mathbb{R}^n$ un ensemble fermé, convexe. Ce problème est équivalent au problème minimiser $f(x)$ sur $\mathbb{R}^n$ où $f = h + I_C$.
Prouvons que F est bien défini, c'est-à-dire, que dom $F = \mathbb{R}^n$.

Proposition 2.1.1 *Pour chaque $x \in I\!R^n$, le problème*

$$\min_{y \in I\!R^n} \{f(y) + \frac{c}{2}\|y - x\|^2\}$$

a une solution unique notée $p_f(x)$. De plus, $p_f(x)$ est l'unique point $y \in I\!R^n$ tel que $c(x - y) \in \partial f(y)$.

Preuve :

Soit $x \in \mathbb{R}^n$ et considérer la fonction

$$\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \quad \tilde{f}(y) = f(y) + \frac{c}{2}\|y - x\|^2.$$

Puisque $\tilde{f}$ est strictement convexe, il existe au plus un minimum.
D'autre part, puisque f est propre, convexe et s.c.i, f est fermée et minorée par une fonction affine. Donc

$$\exists\, s \neq 0\, \exists\, r \in \mathbb{R} \text{ tel que } \forall y \quad \tilde{f}(y) \geq \langle s,\, y\rangle + r + \frac{c}{2}\|y-x\|^2.$$

D'où $\tilde{f}$ est fermée et coercive, ce qui implique que $\tilde{f}$ admet au moins un minimum.
Notons $p_f(x)$ l'unique minimum. En utilisant les conditions d'optimalité, ce minimum est caractérisé par

$$0 \in \partial(f + \frac{c}{2}\|\cdot - x\|^2)(p_f(x))$$

ou encore puisque, $\frac{c}{2}\|\cdot - x\|^2$ a $\mathbb{R}^n$ comme domaine,

$$0 \in \partial f(p_f(x)) + c[p_f(x) - x],$$

c'est-à-dire, $c[x - p_f(x)] \in \partial f(p_f(x))$. □

Le vecteur $p_f(x)$ est appelé le point proximal de x, et satisfait la propriété suivante

$$\forall x \in \mathbb{R}^n \quad F(x) = f(p_f(x)) + \frac{c}{2}\|p_f(x) - x\|^2.$$

Voici quelques propriétés de la régularisée de Moreau-Yosida F.

Théorème 2.1.1 *La régularisée de Moreau-Yosida F est partout définie, convexe et différentiable. Son gradient est donné par*

$$\nabla F(x) = s_f(x) = c[x - p_f(x)] \in \partial(p_f(x)).$$

Sa conjuguée est $F^\star : \mathbb{R}^n \to \mathbb{R} \quad F^\star(s) = f^\star(s) + \frac{1}{2c}\|s\|^2$.
De plus, pour tout x et x' dans $\mathbb{R}^n$,

$$\|\nabla F(x) - \nabla F'(x)\|^2 \leq c\langle \nabla F(x) - \nabla F(x'),\, x - x'\rangle$$

et

$$\|\nabla F(x) - \nabla F'(x)\| \leq c\|x - x'\|.$$

C'est-à-dire $\nabla F(x)$ est Lipschitz sur $\mathbb{R}^n$ de constante c.

Preuve :

1. F est convexe

Soit x_1, $x_2 \in \mathbb{R}^n$ et $t \in]0,1[$. Alors nous avons

$$\begin{aligned} F(tx_1 + (1-t)x_2) &= \min_{y\in\mathbb{R}^n}\{f(y) + \frac{c}{2}\|y - tx_1 - (1-t)x_2\|^2\} \\ F(x_1) &= f(p_f(x_1)) + \frac{c}{2}\|p_f(x_1) - x_1\|^2 \\ F(x_2) &= f(p_f(x_2)) + \frac{c}{2}\|p_f(x_2) - x_2\|^2. \end{aligned}$$

Donc, nous pouvons écrire successivement

$tF(x_1) + (1-t)F(x_2)$

$$\begin{aligned} &= tf(p_f(x_1)) + (1-t)f(p_f(x_2)) + \frac{tc}{2}\|p_f(x_1) - x_1\|^2 + \frac{(1-t)c}{2}\|p_f(x_2) - x_2\|^2 \\ &\geq tf(p_f(x_1)) + (1-t)f(p_f(x_2)) + \frac{c}{2}\|t(p_f(x_1) - x_1) + (1-t)(p_f(x_2) - x_2)\|^2 \\ &= f(\bar{x}) + \frac{c}{2}\|\bar{x} + tx_1 - (1-t)x_2\|^2 \\ &\geq \min_{y\in\mathbb{R}^n}\{f(y) + \frac{c}{2}\|y + tx_1 - (1-t)x_2\|^2\} \\ &= F(tx_1 + (1-t)x_2), \end{aligned}$$

où $\bar{x} = tp_f(x_1) + (1-t)p_f(x_2)$.

Mais cela signifie que F est convexe.

2. F est différentiable et $\nabla F(x) = s_f(x) = c[x - p_f(x)] \in \partial(p_f(x))$.

Il suffit de prouver que $F'(x;\, d) = c[x - p_f(x)]^T d$ pour tout $d \in R^n$.

Soit $d \in \mathbb{R}^n$, $d \neq 0$ et $t > 0$. Alors

$$\begin{aligned} \frac{F(x+td) - F(x)}{t} &\leq \frac{f(p_f(x)) + (c/2)\|p_f(x) - x - td\|^2 - F(x)}{t} \\ &= \frac{f(p_f(x)) + (c/2)\|p_f(x) - x\|^2 + (ct^2/2)\|d\|^2 + ct\langle x - p_f(x),\, d\rangle - F(x)}{t} \\ &= (ct/2)\|d\|^2 + c\langle x - p_f(x),\, d\rangle. \end{aligned}$$

D'où $F'(x;\, d) \leq c\langle x - p_f(x),\, d\rangle$ pour tout $d \in \mathbb{R}^n$. Puisque $F'(x;\, \cdot)$ est convexe, nous avons

$$0 = F'(x;\, \frac{1}{2}d + \frac{1}{2}(-d)) \leq \frac{1}{2}F'(x;\, d) + \frac{1}{2}F'(x;\, -d)$$

et donc,

$$F'(x;\, d) \geq -F'(x;\, -d) \geq c\langle x - p_f(x),\, d\rangle.$$

Par conséquent, $F'(x;\, d) = c[x - p_f(x)]^T d$ pour tout $d \in \mathbb{R}^n$. Donc $\nabla F(x) = s_f(x) = c[x - p_f(x)]$ et par la Proposition 2.1.1, $\nabla F(x) \in \partial f(p_f(x))$.

3. Calcul de $F^\star$

Voir [7] Théorème 6.5.4

4. ∇F est Lipschitz

Soit x, $x' \in \mathbb{R}^n$. Alors, par la seconde partie du Théorème, nous avons que

$$\nabla F(x) - \nabla F(x') = c(x - x') - c[p_f(x) - p_f(x')].$$

Prenant le produit scalaire au deux membres de l'égalité avec $\nabla F(x) - \nabla F(x')$, nous obtenons
$\|\nabla F(x) - \nabla F(x')\|^2 =$

$$c\langle \nabla F(x) - \nabla F(x'),\, x - x'\rangle - c\langle \nabla F(x) - \nabla F(x'),\, p_f(x) - p_f(x')\rangle.$$

Puisque $\nabla F(x) \in \partial f(p_f(x))$, $\nabla F(x') \in \partial f(p_f(x'))$ et ∂f est un opérateur monotone, nous avons

$$\langle \nabla F(x) - \nabla F(x'),\, p_f(x) - p_f(x')\rangle \geq 0.$$

Donc, utilisant l'inégalité de Cauchy-Schwartz, nous obtenons finalement que
$\|\nabla F(x) - \nabla F(x')\|^2 \leq$

$$c\langle \nabla F(x) - \nabla F(x'),\, x - x'\rangle \leq c\|\nabla F(x) - \nabla F(x')\|\|x - x'\|,$$

c'est-à-dire $\|\nabla F(x) - \nabla F(x')\| \leq c\|x - x'\|$. □

Lemme 2.1.1 *Pour tout $x \in \mathbb{R}^n$,*

$$F(x) + (1/2c)\|s_f(x)\|^2 = f(p_f(x)) + (1/c)\|s_f(x)\|^2 \leq f(x).$$

Preuve :

Puisque $F(x) = f(p_f(x)) + (c/2)\|p_f(x) - x\|^2$ et puisque $s_f(x) = c[x - p_f(x)]$, nous avons

$$F(x) + (1/2c)\|s_f(x)\|^2 = f(p_f(x)) + (1/c)\|s_f(x)\|^2.$$

D'autre part, puisque $s_f(x) \in \partial f(p_f(x))$, nous avons

$$f(x) \geq f(p_f(x)) + \langle s_f(x),\, x - p_f(x)\rangle = f(p_f(x)) + (1/c)\|s_f(x)\|^2.$$

□

Théorème 2.1.2

(1) $\inf_{x\in I\!R^n} F(x) = \inf_{x\in I\!R^n} f(x)$

(2) Les points suivants sont équivalents :

(i) x *minimise* f,
(ii) $p_f(x) = x$,
(iii) $s_f(x) = 0$,
(iv) x *minimise* F,
(v) $f(p_f(x)) = f(x)$,
(vi) $F(x) = f(x)$.

Preuve :

(1) Du Théorème 2.1.1, nous avons $F^\star(0) = f^\star(0)$, c'est-à-dire,

$$\sup_{x\in\mathbb{R}^n} \{\langle 0, x\rangle - F(x)\} = \sup_{x\in\mathbb{R}^n} \{\langle 0, x\rangle - f(x)\}.$$

Mais cela signifie que $\inf_{x\in\mathbb{R}^n} F(x) = \inf_{x\in\mathbb{R}^n} f(x)$.

(2) (i) $\Rightarrow$ (ii). Si x minimise f, alors $y = x$ minimise la fonction $\tilde{f}(y) = f(y) + (c/2)\|y - x\|^2$ et donc $x = p_f(x)$ parce que $p_f(x)$ est l'unique minimum de $\tilde{f}$.
(ii) $\Leftrightarrow$ (iii) parce que $s_f(x) = c[x - p_f(x)]$.
(iii) $\Leftrightarrow$ (iv) parce que $\nabla F(x) = s_f(x)$.
(iv) $\Rightarrow$ (v) parce que (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (v).
(v) $\Rightarrow$ (iii) $\Rightarrow$ (iv). Par le Lemme précédent nous avons (v) $\Rightarrow \|s_f(x)\|^2 \leq 0$ et donc (iii) et alors (iv).
(v) $\Rightarrow$ (vi). Puisque (v) $\Rightarrow$ (iii), il suit du Lemme précédent que $F(x) = f(p_f(x))$ et donc, utilisant (v), $F(x) = f(x)$.
(vi) $\Rightarrow$ (i). Toujours par le Lemme précédent nous avons (vi) $\Rightarrow \|s_f(x)\|^2 \leq 0$ et donc (iii) et (ii). D'où $0 = s_f(x) \in \partial f(p_f(x)) = \partial f(x)$. $\square$

2.2 L'algorithme du point proximal.

Par le Théorème 2.1.2, minimiser f est équivalent à trouver un point fixe de l'opérateur-prox p_f.

Algorithme 2.2.1
Pas 1 : Choisir $x^0 \in I\!R^n$ *et* $t_0 > 0$. *Poser* $k = 0$.

Pas 2 : Calculer $x^{k+1} = p_f(x^k)$ en résolvant le problème

$$\min_{y \in I\!R^n} \{f(y) + \frac{1}{2t_k}\|y - x^k\|^2\}$$

Pas 3 : Si $x^{k+1} = x^k$ **on arrête,** *x^{k+1} est un minimum de f.*
Pas 4 : Choisir $t_{k+1} > 0$. Mettre $k = k+1$ et aller au Pas 2.

Quand $t_k = 1/c$ pour tout k, le gradient $\nabla F(x^k)$ est égal à $c(x^k - p_f(x^k))$. Alors $x^{k+1} = p_f(x^k)$ devient $x^{k+1} = x^k - (1/c)\nabla F(x^k)$. Donc, dans ce cas, la méthode du point proximal est équivalente à la méthode du gradient avec une longueur de pas $1/c$.

2.3 Convergence de la méthode

Soit la suite $\{x^k\}$ générée par la méthode du point proximal. Par la définition de $x^{k+1} = \text{argmin}_y\{f(y) + \dfrac{1}{2t_k}\|y - x^k\|^2\}$, nous avons

$$\gamma^k \equiv (1/t_k)(x^k - x^{k+1}) \in \partial f(x^{k+1}).$$

Donc la prox-itération peut être écrite comme $x^{k+1} = x^k - t_k\gamma^k$ avec $\gamma^k \in \partial f(x^{k+1})$. Avant de pouvoir démontrer la convergence de cette méthode nous avons besoin de deux lemmes.

Lemme 2.3.1 *Soit $\bar{z}$ un point limite de la suite $\{z^k\}$ satisfaisant*

$$\|z^{k+1} - \bar{z}\|^2 \leq \|z^k - \bar{z}\|^2 + \delta_k$$

où $\{\delta_k\}$ est une suite de nombres positifs tels que $\sum_{k=0}^{+\infty} \delta_k < +\infty$. Alors la suite converge vers $\bar{z}$.

Preuve :

Soit $\eta > 0$. Alors il existe k_1 tel que

$$\|z^{k_1} - \bar{z}\| \leq \eta/2 \text{ et } \sum_{i=k_1}^{+\infty} \delta_i \leq \eta^2/2$$

et, pour tout $k \geq k_1$,

$$\|z^{k+1} - \bar{z}\|^2 \leq \|z^{k_1} - \bar{z}\|^2 + \sum_{i=k_1}^{k} \delta_i \leq \eta^2/4 + \eta^2/2 \leq \eta^2.$$

Ce qui signifie que $z^k \to \bar{z}$. $\square$

Lemme 2.3.2 *Soit $\{x^k\}$ une suite générée par la méthode du point proximal. Alors*

(i) $\gamma_k \in \partial_{\epsilon_k} f(x^k)$ *avec*

$$\epsilon_k = f(x^k) - f(x^{k+1}) - t_k\|\gamma^k\|^2 \geq 0. \tag{2.1}$$

(ii) La suite $\{f(x^k)\}$ est décroissante.

(iii) Si f est bornée inférieurement, on a que $\epsilon_k \to 0$ et $t_k\|\gamma^k\|^2 \to 0$.

(iv) Pour tout $y \in I\!R^n$ et $k \in I\!N$

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 + t_k^2\|\gamma^k\|^2 + 2t_k[f(y) - f(x^k) + \epsilon_k].$$

Preuve :

(i) Premièrement, observons que $\epsilon_k \geq 0$. En effet, puisque $\gamma^k = (1/t_k)(x^k - x^{k+1}) \in \partial f(x^{k+1})$, nous avons

$$f(x^k) \geq f(x^{k+1}) + \langle \gamma^k, x^k - x^{k+1} \rangle = f(x^{k+1}) + t_k\|\gamma^k\|^2$$

et donc $\epsilon_k \geq 0$. Donc, par la définition de ϵ_k, nous avons

$$f(x^{k+1}) = f(x^k) - t_k\|\gamma^k\|^2 - \epsilon_k = f(x^k) + \langle \gamma^k, x^{k+1} - x^k \rangle - \epsilon_k,$$

et, puisque $\gamma^k \in \partial(x^{k+1})$, nous obtenons que $\gamma^k \in \partial_{\epsilon_k} f(x^k)$ grâce à la formule de transport (voir la Proposition 1.3.2).

(ii) Puisque $\epsilon_k \geq 0$, nous avons, par (2.1), que $f(x^k) \geq f(x^{k+1})$. Donc, la suite $\{f(x^k)\}$ est décroissante.

(iii) Comme f est bornée inférieurement, la suite $f(x^k) \to \bar{f} \in I\!R$. Alors

$$\epsilon_k + t_k\|\gamma^k\|^2 = f(x^k) - f(x^{k+1}) \to 0$$

quand $k \to +\infty$. En particulier, ϵ_k et $t_k\|\gamma^k\|^2$ convergent vers zéro.

(iv) Soit $y \in I\!R^n$ et $k \in I\!N$. Nous avons successivement

$$\begin{aligned}\|x^{k+1} - y\|^2 &= \|x^{k+1} - x^k + x^k - y\|^2 \\ &= \|x^{k+1} - x^k\|^2 + 2\langle x^{k+1} - x^k, x^k - y \rangle + \|x^k - y\|^2 \\ &= t_k^2\|\gamma^k\|^2 - 2\langle t_k\gamma^k, x^k - y \rangle + \|x^k - y\|^2. \end{aligned} \tag{2.2}$$

Puisque $\gamma^k \in \partial_{\epsilon_k} f(x^k)$, nous avons

$$f(y) \geq f(x^k) + \langle \gamma^k, y - x^k \rangle - \epsilon_k$$

c'est-à-dire,

$$\langle \gamma^k, y - x^k \rangle \leq f(y) - f(x^k) + \epsilon_k.$$

Utilisant cette inégalité dans (2.2) , nous obtenons

$$\|x^{k+1} - y\|^2 \leq t_k^2 \|\gamma^k\|^2 + 2t_k[f(y) - f(x^k) + \epsilon_k] + \|x^k - y\|^2.$$

□

Théorème 2.3.1 *Soit $\{x^k\}$ une suite générée par l'algorithme du point proximal. Si $\sum_{k=0}^{+\infty} t_k = +\infty$, alors*

(i) $\lim_{k \to +\infty} f(x^k) = f^\star = \inf_{x \in I\!R^n} f(x)$,

(ii) la suite $\{x^k\}$ converge vers le minimum de f (si il y en a un).

Preuve :

(i) Puisque la suite $\{f(x^k)\}$ est décroissante, elle converge vers un $\bar{f}$. Supposons maintenant, pour obtenir une contradiction, que $\bar{f} > f^\star = \inf_{x \in \mathbb{R}^n} f(x)$. Alors, il existe $y \in \mathbb{R}^n$ et $\delta > 0$ tel que pour tout k, $f(y) + \delta < f(x^k)$. Puisque $t_k \|\gamma^k\|^2 + \epsilon_k \to 0$, il existe k_0 tel que, pour $k \geq k_0$, $t_k \|\gamma^k\|^2 + \epsilon_k < \delta/2$.
Alors, par le Lemme 2.3.2 partie (iv), nous avons, pour tout $k \geq k_0$,

$$\begin{aligned} \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 + t_k[2t_k\|\gamma^k\|^2 + 2\{f(y) - f(x^k) + \epsilon_k\}] \\ &\leq \|x^k - y\|^2 + t_k[\delta - 2\delta] = \|x^k - y\|^2 - \delta t_k. \end{aligned}$$

En sommant, nous obtenons, pour tout $k > k_0$,

$$0 \leq \|x^k - y\|^2 \leq \|x^{k_0} - y\|^2 - \delta \sum_{i=k_0}^{k-1} t_i$$

et passant à la limite sur k, nous avons $\sum_{i=k_0}^{+\infty} t_i \leq \|x^{k_0} - y\|^2$ ce qui contredit l'hypothèse.

(ii) Soit $x^\star$ le minimum de f. Alors par le Lemme 2.3.2 partie (iv), nous avons

$$\begin{aligned} \|x^{k+1} - x^\star\|^2 &\leq \|x^k - x^\star\|^2 + t_k^2\|\gamma^k\|^2 + 2t_k[f(x^\star) - f(x^k) + \epsilon_k] \\ &\leq \|x^k - x^\star\|^2 + t_k^2\|\gamma^k\|^2 + 2t_k[f(x^{k+1}) - f(x^k) + \epsilon_k]. \end{aligned}$$

Par la définition de ϵ_k, $2t_k[f(x^{k+1}) - f(x^k) + \epsilon_k] = -2t_k^2\|\gamma^k\|^2$, et donc $\|x^{k+1} - x^\star\|^2 \leq \|x^k - x^\star\|^2$. En particulier, cela signifie que la suite $\{x^k\}$ est bornée. Soit $\bar{x}$ un point limite de $\{x^k\}$. Alors $\{x^{k_j}\} \to \bar{x}$ et, par la première partie

$$f(\bar{x}) = \lim f(x_{k_j}) = \inf_{x \in \mathbb{R}^n} f(x).$$

Donc $\bar{x}$ est un minimum de f. Finalement, puisque $\|x^{k+1} - \bar{x}\| \leq \|x^k - \bar{x}\|$, il suit par le Lemme 2.3.1 que la suite converge vers $\bar{x}$. $\square$

2.4 La méthode du point proximal approximée

Bien souvent, le problème de trouver $p_f(x^k)$, c'est-à-dire, de résoudre

$$\min_{y \in \mathbb{R}^n} \{f(y) + (1/2t_k)\|y - x^k\|^2\} \tag{2.3}$$

est aussi difficile que la résolution du problème initial. Nous allons présenter une implémentation de ce pas basée sur l'hypothèse qu'en x^k, seules la valeur $f(x^k)$ et le sous-gradient $s(x^k) \in \partial f(x^k)$ sont disponibles grâce à un "oracle." Dans cette section, nous supposons que la fonction est à valeur finie et convexe. La stratégie est la suivante : à chaque itération, la fonction f est remplacée par une fonction φ^k convexe plus simple et pour laquelle le problème

$$\min_{y \in \mathbb{R}^n} \{\varphi^k(y) + (1/2t_k)\|y - x^k\|^2\} \tag{2.4}$$

est plus facile à résoudre et la convergence est préservée. Par exemple, si la fonction φ^k est continue linéaire par morceaux de la forme $\varphi^k(x) = \max_{1 \leq j \leq m}\{a_j^T x + b_j\}$ où, pour tout j, $a_j \in \mathbb{R}^n$ et $b_j \in \mathbb{R}$, alors le sous-problème (2.4) est équivalent au problème quadratique convexe

$$\begin{cases} \min & v + (1/2t_k)\|y - x^k\|^2 \\ \text{s.c} & a_j^T y + b_j \leq v. \end{cases}$$

Il existe beaucoup de méthodes efficaces pour résoudre ce problème.
Maintenant soit y^{k+1} la solution du problème (2.4). Si la décroissance $f(x^k) - f(y^{k+1})$ est suffisante, c'est-à-dire, si

$$f(x^k) - f(y^{k+1}) \geq m[f(x^k) - \varphi^k(y^{k+1})], \tag{2.5}$$

où $0 < m < 1$, alors x^{k+1} est mis à y^{k+1}. Sinon, le modèle de la fonction φ^k est " amélioré". La condition (2.5) est similaire à celle utilisée pour les régions de confiance. La décroissance réelle est comparée avec la décroissance prédite du modèle, à savoir : $f(x^k) - \varphi^k(y^{k+1})$. Nous allons introduire la définition suivante.

Définition 2.4.1 *Soit $m \in (0,1)$ et $x^k \in \mathbb{R}^n$. Une fonction convexe φ^k est une m-approximation de f en x^k si $\varphi^k \leq f$ et*

$$f(x^k) - f(y^{k+1}) \geq m[f(x^k) - \varphi^k(y^{k+1})],$$

où $y^{k+1} = argmin\{\varphi^k(y) + (1/2t_k)\|y - x^k\|^2\}$.

2.4.1 Un algorithme général

Algorithme 2.4.1 *Soit $m \in (0,1)$ et $\{t_k\}$ une suite de nombres positifs. Choisir un point de départ x^0 et poser $k = 0$.*

Pas 1. Trouver φ^k une m-approximation de f en x^k et noter x^{k+1} l'unique solution du problème (2.4).

Pas 2. Augmenter k d'une unité et aller au Pas 1.

Par optimalité de x^{k+1}, nous avons $0 \in \partial\varphi^k(x^{k+1}) + (1/t_k)(x^{k+1} - x^k)$, c'est-à-dire,

$$\gamma^k := \frac{1}{t_k}(x^k - x^{k+1}) \in \partial\varphi^k(x^{k+1}).$$

Donc, la décroissance prédite par le modèle φ^k, à savoir $f(x^k) - \varphi^k(x^{k+1})$, est aussi positive. En effet, puisque $f \geq \varphi^k$ et $\gamma^k \in \partial\varphi^k(x^{k+1})$, nous avons

$$f(x^k) - \varphi^k(x^{k+1}) \geq \varphi^k(x^k) - \varphi^k(x^{k+1}) \geq \langle \gamma^k, x^k - x^{k+1} \rangle = t_k\|\gamma^k\|^2 > 0$$

si $x^{k+1} \neq x^k$.

Nous allons maintenant prouver la convergence de l'algorithme général.

Théorème 2.4.1 *Soit $\{x^k\}$ une suite générée par l'algorithme général. Si $\sum_{k=1}^{+\infty} t_k = +\infty$, alors $f(x^k) \searrow \bar{f} = \inf_x f(x)$.*
Si en plus $t_k \leq \bar{t}$ pour tout k, alors $x^k \to x^*$ où x^* est un minimum de f (si un tel minimum existe).

Preuve :

Par définition, nous avons $\gamma^k = (x^k - x^{k+1})/(t_k)$ et donc $x^{k+1} = x^k - t_k\gamma^k$. Avec ces notations, nous pouvons prouver que :

a. $\{f(x^k)\}$ est décroissante.
Puisque la décroissance prédite par le modéle φ^k est positive, nous avons

$$f(x^k) - f(x^{k+1}) \geq m[f(x^k) - \varphi^k(x^{k+1})] \geq 0, \qquad (2.6)$$

et donc $\{f(x^k)\}$ est décroissante. Pour la suite, nous supposons que $\{f(x^k)\}$ est bornée inférieurement (sinon $f(x^k) \searrow -\infty$ et la preuve est finie).

b. $\gamma^k \in \partial_{\epsilon_k} f(x^k)$ où $\epsilon_k = f(x^k) - \varphi^k(x^{k+1}) - t_k\|\gamma^k\|^2$.
Pour chaque y et k, nous avons succesivement

$$\begin{aligned} f(y) \geq \varphi^k(y) &\geq \varphi^k(x^{k+1}) + \langle \gamma^k,\, y - x^{k+1}\rangle \\ &= f(x^k) + \varphi^k(x^{k+1}) - f(x^k) + \langle \gamma^k,\, y - x^k\rangle + \langle \gamma^k,\, x^k - x^{k+1}\rangle \\ &= f(x^k) + \varphi^k(x^{k+1}) - f(x^k) + \langle \gamma^k,\, y - x^k\rangle + t_k\|\gamma^k\|^2. \end{aligned}$$

Donc $\gamma^k \in \partial_{\epsilon_k} f(x^k)$ avec $\epsilon_k = f(x^k) - \varphi^k(x^{k+1}) - t_k\|\gamma^k\|^2$.

c. $\sum_{k=1}^{+\infty}\{\epsilon_k + t_k\|\gamma^k\|^2\} < +\infty$.
Par (2.6), nous avons

$$\epsilon_k = f(x^k) - \varphi^k(x^{k+1}) - t_k\|\gamma^k\|^2 \leq (1/m)(f(x^k) - f(x^{k+1})) - t_k\|\gamma^k\|^2.$$

Donc

$$\sum_{k=1}^{n}\{\epsilon_k + t_k\|\gamma^k\|^2\} \leq (1/m)\sum_{k=1}^{n}\{f(x^k) - f(x^{k+1})\} = (1/m)\{f(x^1) - f(x^{n+1})\}.$$

Puisque $\{f(x^k)\}$ est bornée inférieurement, la série $\sum_{k=1}^{+\infty}\{\epsilon_k + t_k\|\gamma^k\|^2\}$ est convergente.

d. $f(x^k) \searrow \bar{f} = \inf_x f(x)$.
Puisque la suite $\{f(x^k)\}$ est décroissante, elle converge vers un certain $\bar{f}$. Supposons maintenant, pour obtenir une contradiction, que $\bar{f} > f^\star = \inf_{x\in\mathbb{R}^n} f(x)$. Alors il existe $y \in \mathbb{R}^n$ et $\delta > 0$ tel que pour tout k, $f(y)+\delta < f(x^k)$. Puisque $t_k\|\gamma^k\|^2 + \epsilon_k \to 0$, il existe k_0 tel que, pour $k \geq k_0$, $t_k\|\gamma^k\|^2 + \epsilon_k < \delta/2$. Alors, par le Lemme 2.3.2 (iv), nous avons

$$\begin{aligned} \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 + t_k[2t_k\|\gamma^k\|^2 + 2\{f(y) - f(x^k) + \epsilon_k\}] \\ &\leq \|x^k - y\|^2 + t_k[\delta - 2\delta] = \|x^k - y\|^2 - \delta t_k. \end{aligned}$$

En sommant, nous obtenons, pour tout $k > k_0$,

$$0 \leq \|x^k - y\|^2 \leq \|x^{k_0} - y\|^2 - \delta\sum_{i=k_0}^{k-1} t_i$$

et en passant à la limite sur k, nous avons $\sum_{i=k_0}^{+\infty} t_i \leq \dfrac{1}{\delta}\|x^{k_0} - y\|^2$ ce qui contredit l'hypothèse.

Maintenant, supposons que la fonction f a un minimum en $\bar{x}$ et que la suite $\{t_k\}$ est bornée.

e. $\{x^k\}$ est bornée.
Utilisant le Lemme 2.3.2 (iv) avec $y = \bar{x}$, nous obtenons

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + t_k^2\|\gamma^k\|^2 + 2t_k[f(\bar{x}) - f(x^k) + \epsilon_k].$$

Puisque $f(\bar{x}) - f(x^k) \leq 0$ et $t_k \leq \bar{t}$, nous avons

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + 2\bar{t}[t_k\|\gamma^k\|^2 + \epsilon_k]$$

et, utilisant la partie c. de cette preuve,

$$\sum_{k=1}^{+\infty} \bar{t}[t_k\|\gamma^k\|^2 + \epsilon_k] < +\infty.$$

D'où la suite $\{x^k - \bar{x}\}_k$ est bornée et donc aussi la suite $\{x^k\}$.

Soit $x^\star$ un point limite de $\{x^k\}$.

f. $x^\star$ est un minimum de f et $x^k \to x^\star$.
Puisque $x^{n_k} \to x^\star$, nous avons, par contuinité de f, que $f(x^{n_k}) \to f(x^\star)$. Alors, utilisant la partie d. de la preuve, nous pouvons conclure que $f(x^\star) = \bar{f}$, c'est-à-dire, $x^\star$ est un minimum de f. Utilisant aussi le Lemme 2.3.2 (iv) mais avec $y = x^\star$, nous obtenons comme ci-dessus que

$$\|x^{k+1} - x^\star\|^2 \leq \|x^k - x^\star\|^2 + 2t_k[t_k\|\gamma^k\|^2 + \epsilon_k].$$

Puisque $x^\star$ est un point limite de la suite $\{x^k\}$ et puisque la série $\sum_{k=1}^{+\infty} t_k[t_k\|\gamma^k\|^2 + \epsilon_k]$ est convergente, nous pouvons utiliser le Lemme 2.3.1 et nous obtenons que la suite $\{x_k\}$ converge vers $x^\star$. □

2.4.2 Construction des $m-$approximations

Pour obtenir un algorithme implémentable, nous devons maintenant dire comment calculer les $m-$approximations φ^k de f en x^k de telle façon que le sous-problème (2.4) est plus facile à résoudre que le problème (2.3). Nous savons déjà que si φ^k est convexe et linéaire par morceaux, alors le sous-problème (2.4) est équivalent à un problème quadratique convexe et il existe des méthodes numériques efficaces pour résoudre ce problème. Donc, si nous voulons obtenir une fonction convexe linéaire par morceaux pour la fonction modèle φ^k, il est judicieux de la construire morceaux par morceaux en générant successivement

$$\varphi_i^k,\ i = 1, 2, \ldots$$

jusqu'a ce que (si possible) $\varphi_{i_k}^k$ est une $m-$approximation de f en x^k pour un $i_k \geq 1$. Pour $i = 1, 2, \ldots$, nous notons par y_i^k l'unique solution du problème

$$(P_i^k) \qquad \min_y\{\varphi_i^k(y) + (1/2t_k)\|y - x^k\|^2\},$$

et nous posons $\varphi^k = \varphi^k_{i_k}$ et $x^{k+1} = y^k_{i_k}$.

Dans l'ordre pour obtenir une $m-$approximation $\varphi^k_{i_k}$ de f en x^k, nous devons imposer quelques conditions sur les modèles successifs φ^k_i, $i = 1, 2, \ldots$. Cependant, avant de les présenter, nous devons définir les fonctions affines l^k_i, $i = 1, 2, \ldots$ par

$$l^k_i(y) = \varphi^k_i(y^k_i) + \langle \gamma^k_i, y - y^k_i \rangle \quad \forall y \in \mathbb{R}^n,$$

où $\gamma^k_i = \frac{1}{t_k}(x^k - y^k_i)$. Par les conditions d'optimalité pour y^k_i, nous avons

$$\gamma^k_i \in \partial \varphi^k_i(y^k_i). \tag{2.7}$$

Nous pouvons alors observer que

$$l^k_i(y^k_i) = \varphi^k_i(y^k_i) \text{ et } l^k_i(y) \leq \varphi^k_i(y) \text{ pour tout } y \in \mathbb{R}^n. \tag{2.8}$$

Maintenant, supposons que les conditions suivantes sont satisfaites pour le modèle convexe φ^k_i,

(C1) $\varphi^k_i \leq f$ sur $\mathbb{R}^n$ pour $i = 1, 2, \ldots$

(C2) $\varphi^k_{i+1} \geq f(y^k_i) + \langle s(y^k_i), \cdot - y^k_i \rangle$ sur $\mathbb{R}^n$ pour $i = 1, 2, \ldots$

(C3) $\varphi^k_{i+1} \geq l^k_i$ sur $\mathbb{R}^n$ pour $i = 1, 2, \ldots$

où $s(y^k_i)$ est le sous-gradient de f en y^k_i.

Plusieurs modèles vérifient ces conditions. Par exemple, pour le premier modèle φ^k_1, nous pouvons prendre la fonction linéaire

$$\varphi^k_1(y) = f(x^k) + \langle s(x^k), y - x^k \rangle \quad \forall y \in \mathbb{R}^n.$$

Puisque $s(x^k) \in \partial f(x^k)$, la condition (C1) est vérifiée pour $i = 1$. Pour les modèles suivant $i = 2, \ldots$, il existe différentes possibilités. Nous pouvons prendre par exemple pour $i = 1, 2, \ldots$

$$\varphi^k_{i+1} = \max\{l^k_i(y), f(y^k_i) + \langle s(y^k_i), y - y^k_i \rangle\} \quad \forall y \in \mathbb{R}^n. \tag{2.9}$$

Les conditions (C2) et (C3) sont évidemment satisfaites et la condition (C1) est aussi vérifiée parce que chaque morceau linéaire de ces fonctions est inférieure à f.

Maintenant décrivons l'algorithme permettant de passer de x^k à x^{k+1}.

Algorithme 2.4.2 L'algorithme de pas sérieux.
Soit $x^k \in \mathbb{R}^n$ *et* $m \in (0, 1)$. *Soit* $i = 1$.

Pas 1 : Choisir φ_i^k une fonction convexe qui vérifie (C1) à (C3) et résoudre le problème (P_i^k) pour obtenir y_i^k.

Pas 2 : Si $f(x^k) - f(y_i^k) \geq m[f(x^k) - \varphi_i^k(y_i^k)]$, alors poser $x^{k+1} = y_i^k$, $i_k = i$ et STOP ; x^{k+1} est un pas sérieux.

Pas 3 : Augmenter i d'une unité et aller au Pas 1.

Précisons que φ_i^k satisfait (C1)-(C3) signifie que φ_i^k satisfait (C1) et, si $i \geq 2$, φ_i^k satisfait (C2) et (C3) avec $i+1$ remplacé par i.

Notre but maintenant est de prouver que si x^k n'est pas un minimum de f et si les modèles φ_i^k, $i = 1, \ldots$ satisfont (C1)-(C3), alors il existe $i_k \in \mathbb{N}_0$ tel que $\varphi_{i_k}^k$ est une $m-$approximation de f en x, c'est-à-dire que le STOP à lieu au Pas 2 après un nombre fini d'itérations.
Pour obtenir ce résultat, nous commençons par prouver la proposition suivante.

Proposition 2.4.1 *Supposons que les modèles φ_i^k, $i = 1, 2, \ldots$ vérifient les conditions (C1)-(C3), et soit, pour chaque i, y_i^k est l'unique solution du problème (P_i^k). Alors*
(1) $f(y_i^k) - \varphi_i^k(y_i^k) \to 0$,
(2) $y_i^k \to p_f(x^k)$,
quand $i \to +\infty$.

Avant de donner la preuve de cette proposition, nous introduisons les notations suivantes :

$$\tilde{\varphi}_i^k(y) = \varphi_i^k(y) + (1/2t_k)\|y - x^k\|^2,$$
$$\tilde{l}_i^k(y) = l_i^k(y) + (1/2t_k)\|y - x^k\|^2.$$

Alors nous avons que $\tilde{\varphi}_i^k(x^k) = \varphi_i^k(x^k)$, $\tilde{\varphi}_i^k(y_i^k) = \tilde{l}_i^k(y_i^k)$ et le Lemme :

Lemme 2.4.1 $\tilde{l}_i^k(y) = \tilde{l}_i^k(y_i^k) + (1/2t_k)\|y - y_i^k\|^2$ *pour tout i et y.*

Preuve :

Utilisant la définition de $\tilde{l}_i^k$, l_i^k et γ_i^k, nous avons

$$\tilde{l}_i^k(y_i^k) + (1/2t_k)\|y - y_i^k\|^2 = l_i^k(y_i^k) + (1/2t_k)\|y_i^k - x^k\|^2 + (1/2t_k)\|y - y_i^k\|^2$$

$$
\begin{aligned}
&= \varphi_i^k(y_i^k) + (1/2t_k)\|y - x^k\|^2 - (1/t_k)\langle y_i^k - x^k,\, y - y_i^k\rangle \\
&= \varphi_i^k(y_i^k) + (1/2t_k)\|y - x^k\|^2 + \langle \gamma_i^k,\, y - y_i^k\rangle \\
&= l_i^k(y) + (1/2t_k)\|y - x^k\|^2 = \tilde{l}_i^k(y).
\end{aligned}
$$

□

Preuve de la proposition :

1. $\{\tilde{l}_i^k(y_i^k)\}$ est convergente et $y_{i+1}^k - y_i^k \to 0$ quand $i \to +\infty$.

Pour tout $i \geq 1$, nous avons

$$
\begin{array}{lll}
f(x^k) & \geq \varphi_{i+1}^k(x^k) = \tilde{\varphi}_{i+1}^k(x^k) & \\
 & \geq \tilde{\varphi}_{i+1}^k(y_{i+1}^k) & \text{(définition de } y_{i+1}^k) \\
 & = \tilde{l}_{i+1}^k(y_{i+1}^k) & \text{(définition de } l_{i+1}^k) \\
 & \geq \tilde{l}_i^k(y_{i+1}^k) & \text{(parce que } \varphi_{i+1}^k \geq l_i^k) \\
 & = \tilde{l}_i^k(y_i^k) + (1/2t_k)\|y_{i+1}^k - y_i^k\|^2. & \text{(Par le Lemme 2.4.1 avec } y = y_{i+1}^k).
\end{array}
$$

Avec ces relations, nous pouvons déduire que

- la suite $\{\tilde{l}_i^k(y_i^k)\}$ est croissante et bornée supérieurement par $f(x^k)$; donc elle est convergente.
- $\tilde{l}_{i+1}^k(y_{i+1}^k) - \tilde{l}_i^k(y_i^k) \geq (1/2t_k)\|y_{i+1}^k - y_i^k\|^2 \geq 0$, ce qui implique que $y_{i+1}^k - y_i^k \to 0$ parce que le coté de droit de la dernière inégalité tend vers zéro.

2. La suite $\{y_i^k\}_i$ est bornée.

Soit $y \in \mathbb{R}^n$. Alors, puisque $f \geq l_i^k$, il suit, du Lemme 2.4.1, que

$$
f(y) + (1/2t_k)\|y - x^k\|^2 \geq \tilde{l}_i^k(y) = \tilde{l}_i^k(y_i^k) + (1/2t_k)\|y - y_i^k\|^2.
$$

Puisque $\{\tilde{l}_i^k(y_i^k)\}$ est convergente, la suite $\{y - y_i^k\}$ doit être bornée et donc la suite $\{y_i^k\}$ aussi.

3. $f(y_i^k) - \varphi_i^k(y_i^k) \to 0$ quand $i \to +\infty$.

Nous avons successivement

$$
\begin{array}{lll}
f(y_{i+1}^k) - f(y_i^k) & \geq \varphi_{i+1}^k(y_{i+1}^k) - f(y_i^k) & \text{(parce que } f \geq \varphi_{i+1}^k) \\
 & \geq \langle s(y_i^k),\, y_{i+1}^k - y_i^k\rangle & \text{(Condition (C2))}.
\end{array}
$$

De plus, puisque f est Lipschitz sur un ensemble borné, nous obtenons

$$
|f(y_{i+1}^k) - f(y_i^k)| \leq L\|y_{i+1}^k - y_i^k\|.
$$

Alors $f(y_{i+1}^k) - f(y_i^k) \to 0$ (parce que $y_{i+1}^k - y_i^k \to 0$). D'autre part, la suite $\{s(y_i^k)\}_i$ est bornée car $\{y_i^k\}_i$ est bornée et le sous-différentiel est borné sur un ensemble borné.

Donc, puisque $y_{i+1}^k - y_i^k \to 0$, nous avons $\langle s(y_i^k), y_{i+1}^k - y_i^k\rangle \to 0$. D'où $\varphi_{i+1}^k(y_{i+1}^k) - f(y_i^k) \to 0$ et

$$\varphi_{i+1}^k(y_{i+1}^k) - f(y_{i+1}^k) = \varphi_{i+1}^k(y_{i+1}^k) - f(y_i^k) + f(y_i^k) - f(y_{i+1}^k) \to 0.$$

4. $y_i^k \to p_f(x^k)$ quand $i \to +\infty$.

Puisque la suite $\{y_i^k\}_i$ est bornée, il est suffisant de prouver que chaque point limite $\bar{y}^k$ de $\{y_i^k\}$ est égale à $p_f(x^k)$, c'est-à-dire, en utilisant la caractérisation de $p_f(x^k)$, que $(x^k - \bar{y}^k)/t_k \in \partial f(\bar{y}^k)$.
Nous avons $y_i^k \to \bar{y}^k$ pour $i \in K \subseteq \mathbb{N}$. Donc, pour tout $i \in K$ et tout $y \in \mathbb{R}^n$, il suit de la définition du sous-différentiel que

$$f(y) \geq \varphi_i^k(y) \geq \varphi_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k\rangle. \tag{2.10}$$

Puisque $f(y_i^k) \to f(\bar{y}^k)$ et $\varphi_i^k(y_i^k) - f(y_i^k) \to 0$, nous avons que $\varphi_i^k(y_i^k) \to f(\bar{y}^k)$.
Cependant $\gamma_i^k = (x^k - y_i^k)/t_k \to (x^k - \bar{y}^k)/t_k \equiv \bar{\gamma}$. Donc, en passant à la limite dans (2.10), nous obtenons que pour tout y, $f(y) \geq f(\bar{y}^k) + \langle\bar{\gamma}, y - \bar{y}^k\rangle$, c'est-à-dire, $\bar{\gamma} \in \partial f(\bar{y}^k)$ où encore $(x^k - \bar{y}^k)/t_k \in \partial f(\bar{y}^k)$. □

Théorème 2.4.2 *Si x^k n'est pas un minimum de f, alors l'algorithme de pas sérieux s'arrête après un nombre fini d'itération i_k avec $\varphi_{i_k}^k$ une $m-$approximation de f en x_k et avec $x^{k+1} = y_{i_k}^k$.*

Preuve :

Supposons pour obtenir une contradiction que l'algorithme de pas sérieux ne s'arrête pas. Alors

$$\forall i \geq 1 \quad f(x^k) - f(y_i^k) < m[f(x^k) - \varphi_i^k(y_i^k)]. \tag{2.11}$$

Par la Proposition 2.4.1, nous obtenons que $y_i^k \to p_f(x^k)$ et $\varphi_i^k(y_i^k) - f(y_i^k) \to 0$ quand $i \to +\infty$. D'où passant à la limite dans (2.11), nous avons

$$f(x^k) - f(p_f(x^k)) \leq m[f(x^k) - f(p_f(x^k))],$$

c'est-à-dire, $(1-m)[f(x^k) - f(p_f(x^k))] \leq 0$.
Puisque $1-m > 0$, nous avons $f(x^k) \leq f(p_f(x^k))$, ou encore par la définition de $\tilde{f}$

$$\tilde{f}(x^k) = f(x^k) \leq f(p_f(x^k)) \leq \tilde{f}(p_f(x^k)).$$

Puisque $p_f(x^k)$ est l'unique minimum de $\tilde{f}$, nous obtenons $x^k = p_f(x^k)$. Mais cela signifie que x^k est un minimum de f, ce qui contredit l'hypothèse du Théorème. Donc l'algorithme de pas sérieux s'arrête après un nombre fini d'itérations. □

2.4.3 Une méthode de faisceaux

Insérant l'algorithme du pas sérieux dans le pas 1 de l'algorithme général nous obtenons l'agorithme suivant :

Algorithme de faisceaux I

Algorithme 2.4.3 *Choisir un point initial* $x^0 \in C$, *une tolérance* $m \in (0,1)$ *et une suite positive* $\{t_k\}_{k \in I\!N}$. *Poser* $y_0^0 = x_0$ *et* $k = 0$, $i = 1$.
<u>Pas 1 :</u> Choisir une fonction convexe linéaire par morceaux φ_i^k *satisfaisant (C1)-(C3) et résoudre*

$$(P_i^k) \qquad \min_y \{\varphi_i^k + (1/2t_k)\|y - x^k\|^2\},$$

pour obtenir l'unique solution optimale y_i^k.
<u>Pas 2 :</u> Si

$$f(x^k) - f(y_i^k) \geq m[f(x^k) - \varphi_i^k(y_i^k)], \tag{2.12}$$

alors poser $x^{k+1} = y_i^k$, $y_0^{k+1} = x^{k+1}$, *augmenter* k *d'une unité et poser* $i = 0$.
<u>Pas 3 :</u> Augmenter i *d'une unité et aller au Pas 1.*

Des Théorèmes 2.4.1 et 2.4.2, nous obtenons le résultat de convergence suivant.

Théorème 2.4.3 *Supposons que* $\sum t_k = +\infty$ *et* $t_k \leq \bar{t}$ *pour tout* k. *Si la suite* $\{x^k\}$ *générée par l'algorithme de faisceaux I est infinie, alors* $\{x^k\}$ *converge vers un minimum de* f. *Si après un certain* k *atteint, le critère (2.12) n'est toujours pas satisfait, alors* x^k *est un minimum de* f.

Pour une implémentation pratique, il est nécessaire de définir un critère d'arrêt. Nous rappelons que $\bar{x}$ est un point $\epsilon-$stationnaire si il existe $s \in \partial_\epsilon f(\bar{x})$ avec $\|s\| \leq \epsilon$. Puisque, par optimalité de y_i^k, $\gamma_i^k \in \partial\varphi_i^k(y_i^k)$, il est facile de prouver que

$$\gamma_i^k \in \partial_{\epsilon_i^k} f(y_i^k)$$

où $\epsilon_i^k = f(y_i^k) - \varphi_i^k(y_i^k)$. En effet, pour tout $y \in \mathbb{R}^n$, nous avons

$$\begin{aligned} f(y) &\geq \varphi_i^k(y) \geq \varphi_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k \rangle \\ &= f(y_i^k) + \langle \gamma_i^k, y - y_i^k \rangle - [f(y_i^k) - \varphi_i^k(y_i^k)]. \end{aligned}$$

d'où nous pouvons introduire le critère d'arrêt : si $f(y_i^k) - \varphi_i^k(y_i^k) \leq \epsilon$ et $\|\gamma_i^k\| \leq \epsilon$, alors y_i^k est un point $\epsilon-$stationnaire.
Pour prouver que le critère d'arrêt est satisfait après un nombre fini d'itérations, nous utilisons la Proposition suivante.

Proposition 2.4.2 *Supposons $0 < \underline{t} \leq t_k \leq \bar{t}$ pour tout k. Si la suite $\{x^k\}$ générée par l'algorithme de faisceaux I est infinie, alors $f(y^k_{i_k}) - \varphi^k_{i_k}(y^k_{i_k}) \to 0$ et $\|\gamma^k_{i_k}\| \to 0$ quand $k \to +\infty$. Si la suite $\{x^k\}$ est finie avec k le dernier indice, alors $f(y^k_i) - \varphi^k_i(y^k_i) \to 0$ et $\|\gamma^k_i\| \to 0$ quand $i \to +\infty$.*

Preuve :

Si la suite $\{x^k\}$ est infinie, alors, par le Théorème 2.4.1, $x^k \to x^\star$ et $y^k_{i_k} = x^{k+1}$ pour tout k. Donc $\gamma^k_{i_k} = (1/t_k)(x^k - x^{k+1}) \to 0$ quand $k \to +\infty$ parce que $t_k \geq \underline{t} > 0$ pour tout k. De plus, nous avons

$$\begin{aligned} 0 &\leq f(y^k_{i_k}) - \varphi^k_{i_k}(y^k_{i_k}) \\ &= f(x^{k+1}) - f(x^k) + f(x^k) - \varphi^k_{i_k}(x^{k+1}) \\ &\leq f(x^{k+1}) - f(x^k) + \frac{1}{m}[f(x^k) - f(x^{k+1})], \end{aligned}$$

où le critère d'arrêt (2.12) est utilisé pour obtenir la dernière inégalité. Puisque f est continue et $x^k \to x^\star$, nous obtenons que $f(y^k_{i_k}) - \varphi^k_{i_k}(y^k_{i_k}) \to 0$ quand $k \to +\infty$.

Maintenant, si la suite $\{x^k\}$ est finie avec k le dernier indice, alors, par le Théorème 2.4.1, x^k est un minimum de f et donc $x^k = p_f(x^k)$. De plus, par la Proposition 2.4.1, $f(y^k_i) - \varphi^k_i(y^k_i) \to 0$ et $y^k_i \to p_f(x^k)$ qaund $i \to +\infty$. Donc

$$\gamma^k_i = (1/t_k)(x^k - y^k_i) \to (1/t_k)(x^k - p_f(x^k)) = 0.$$

Ce qui complète la preuve. □

Et donc l'algorithme devient :

Algorithme de faisceaux II.

Algorithme 2.4.4 *Choisir un point initial $x^0 \in C$, une tolérance $m \in (0,1)$, $\epsilon > 0$, et une suite positive $\{t_k\}_{k \in I\!N}$. Poser $y^0_0 = x_0$ et $k = 0$, $i = 1$.*
Pas 1 : Choisir une fonction convexe linéaire par morceaux φ^k_i satisfaisant (C1)-(C3) et résoudre

$$(P^k_i) \qquad \min_y \{\varphi^k_i + (1/2t_k)\|y - x^k\|^2\},$$

pour obtenir l'unique solution optimale y^k_i. Calculer $\gamma^k_i = (x^k - y^k_i)/t_k$. Si $\|\gamma^k_i\| \leq \epsilon$ et $f(y^k_i) - \varphi^k_i(y^k_i) \leq \epsilon$, alors STOP, y^k_i est un point $\epsilon-$stationnaire.
Pas 2 : Si

$$f(x^k) - f(y^k_i) \geq m[f(x^k) - \varphi^k_i(y^k_i)], \tag{2.13}$$

alors poser $x^{k+1} = y_i^k$, $y_0^{k+1} = x^{k+1}$, *augmenter* k *d'une unité et poser* $i = 0$.
Pas 3 : Augmenter i *d'une unité et aller au Pas 1.*

Chapitre 3

La méthode proximale logarithmique-quadratique.

3.1 La fonction φ.

Soit $\varphi : \mathbb{R} \to (-\infty, +\infty]$ une fonction propre convexe fermée.
Le domaine dom $\varphi := \{t : \varphi(t) < +\infty\} \neq \emptyset$ contenu dans $[0, +\infty)$.
Nous supposons que φ satisfait :

(P1) φ est deux fois continûment différentiable sur int(dom φ)$=(0, +\infty)$.

(P2) φ est fortement convexe sur son domaine.

(P3) $\lim_{t \to 0^+} \varphi'(t) = -\infty$.

(P4) $\varphi(1) = \varphi'(1) = 0$ et $\varphi''(1) > 0$.

Nous notons Φ la classe des fonctions satisfaisant (P1) à (P4).

Nous pouvons remarquer que par la forte convexité de φ et la propriété (P4), nous avons $\varphi(t) = 0 \Leftrightarrow t = 1$.

Nous introduisons deux sous classes de Φ comme suit :

$$\Phi_1 := \{\varphi \in \Phi : \exists\, M > 0 \text{ tel que } \varphi''(t) \leq M, \forall t \geq 1\}, \tag{3.1}$$
$$\Phi_2 := \{\varphi \in \Phi : \varphi''(1)(1 - 1/t) \leq \varphi'(t) \leq \varphi''(1)(1 - t), \forall t > 0\}. \tag{3.2}$$

Exemples 3.1.1 *Les fonctions ci-dessous appartiennent à* $\Phi_1 \cap \Phi_2$.

$$\varphi_1(t) = t \log t - t + 1, \;\; dom\ \varphi_1 = [0, +\infty).$$
$$\varphi_2(t) = -\log t - t + 1, \;\; dom\ \varphi_2 = (0, +\infty).$$
$$\varphi_3(t) = 2(\sqrt{t} - 1)^2, \;\; dom\ \varphi_3 = [0, +\infty).$$

Maintenant considérons les fonctions φ suivantes :

$$\varphi(t) := \mu h(t) + \frac{\nu}{2}(t - 1)^2. \tag{3.3}$$

où $\mu > 0$, $\nu \geq 0$ et $h \in \Phi$.
Nous allons aussi définir un paramètre que nous utiliserons plus tard

$$\theta := \frac{\nu + \mu}{2}. \tag{3.4}$$

Lemme 3.1.1 *Soit $\varphi \in \Phi$, alors*

(i) $(\varphi^*)_\infty(-1) = 0$ et $(\varphi^*)_\infty(1) = +\infty$ où $(\varphi^*)_\infty$ est la fonction de récession de φ^*.

(ii) φ^* est fortement convexe, différentiable sur int dom φ^* et croissante.

(iii) int dom $\varphi^* = (-\infty, \eta)$ avec $0 < \eta \leq +\infty$.

Preuve :

(i) Par [8] (preuve du Théorème 13.3, p 116), nous avons que

$$(\varphi^*)_\infty(s) = \sup \{su : u \in \text{ int dom } \varphi\} = \sup \{su : u > 0\}$$

et (i) suit immédiatement.

(ii) Par les propriétés (P1),(P2),(P3) de la classe des fonctions Φ, nous avons que φ est une fonction de Legendre. Alors par [8] (§ 26), la fonction φ^* est aussi une fonction de Legendre et nous avons (dans la preuve du Corolaire 26.3.1, p 254)

$$(\varphi^*)'(y) = (\varphi')^{-1}(y) \; \forall y \in \text{ int dom } \varphi^*.$$

Mais $(\varphi')^{-1}(y) = x$ est équivalent à $y = \varphi'(x)$ ce qui implique que $x > 0$ et (ii) est satisfait.

(iii) Posons $h(t) := \varphi(t) - ts$ et $\eta := \varphi_\infty(1)$.
Puisque $\varphi \in \Phi$, il existe $t > 1$ tel que $\varphi'(t) > 0$, et alors par [8] (Théorème 13.3, p 116), il suit que

$$0 < \eta \leq \infty.$$

Utilisant la définition de la fonction de récession ([8], p.66), nous obtenons

$$h_\infty(u) = \varphi_\infty(u) - us = \begin{cases} (\eta - s)u & \text{si } u > 0 \\ +\infty & \text{si } u < 0 \\ 0 & \text{sinon.} \end{cases}$$

Par [8] (Corollaire 13.3.4(c), p 117), nous avons que $s \in$ int dom φ^* si et seulement si $h_\infty(u) > 0, \forall u \neq 0$, et d'où par le calcul ci-dessus si et seulement si $s < \eta$. $\square$

Lemme 3.1.2 *Soit* $h \in \Phi_2$, $\varphi(t) = \mu h(t) + (\nu/2)(t-1)^2$, *avec* $\nu \geq \mu h"(1) > 0$ *et posons*

$$\theta := (\nu + \mu h"(1))/2.$$

Pour tout $a,\ b \in I\!R^p_{++}$ *et* $c \in I\!R^p_{+}$, *nous avons*

$$\langle c-b, \Phi'(a,b)\rangle \leq \theta(\|c-a\|^2 - \|c-b\|^2)$$

où $\Phi'(a,b) := (a_1\varphi'(b_1/a_1), \cdots, a_p\varphi'(b_p/a_p))^T$

Preuve :

Soit $t_j = b_j/a_j$, puisque $h \in \Phi_2$, pour chaque $t > 0$ nous avons $h'(t) \leq h"(1)(t-1)$, alors nous obtenons après multiplication par $a_jc_j \geq 0$ pour chaque $j = 1, \cdots, p$,

$$c_ja_jh'\left(\frac{b_j}{a_j}\right) \leq h"(1)c_ja_j\left[\frac{b_j}{a_j} - 1\right] = h"(1)c_j(b_j - a_j).$$

De $-h'(t) \leq -h"(1)(1 - 1/t) \quad \forall t > 0$, pour chaque $j = 1, \cdots, p$ nous avons

$$-a_jb_jh'\left(\frac{b_j}{a_j}\right) \leq h"(1)a_j(a_j - b_j),$$

additionnant les deux inégalités et sommant sur $j = 1, \cdots, p$, nous obtenons alors

$$\langle c-b, H'(a,b)\rangle \leq h"(1)\sum_{j=1}^{p} c_j(b_j - a_j) + a_j(a_j - b_j), \tag{3.5}$$

où $H'(a,b) := (a_1h'(b_1/a_1), \ldots, a_ph'(b_p/a_p))^T$.

Maintenant, puisque $\varphi'(t) = \mu h"(1) + \nu(t-1)$ nous obtenons de (3.5),

$$\langle c-b,\ \Phi'(a,b)\rangle \leq \mu h"(1)\langle b-a,\ c-a\rangle + \nu\langle b-a,\ c-b\rangle.$$

Utilisant les deux égalités suivantes,

$$\langle b-a,\ c-a\rangle = \frac{1}{2}(\|c-a\|^2 - \|c-b\|^2 + \|b-a\|^2),$$

$$\langle b-a,\ c-b\rangle = \frac{1}{2}(\|c-a\|^2 - \|c-b\|^2 - \|b-a\|^2),$$

et par la définition de θ, nous obtenons donc,

$$\langle c-b,\ \Phi'(a,b)\rangle \leq \theta\{\|c-a\|^2 - \|c-b\|^2\} - (v-\theta)\|b-a\|^2, \tag{3.6}$$

et puisque $\theta := (\nu + \mu h"(1))/2$, avec $\nu \geq \mu h"(1) > 0$, nous avons $\nu \geq \theta$, et alors l'inégalité est prouvée. $\square$

Parmi toutes les possibilités de choix de noyau $\varphi \in \Phi_1$ ou Φ_2, il en existe un en particulier qui a des propriétés intéressantes pour développer un algorithme qui résout des problèmes d'optimisation.

Soit $\nu > \mu > 0$ des paramètres fixés. Posons

$$h(t) = \begin{cases} t - \log t - 1 & \text{si } t > 0, \\ +\infty & \text{sinon.} \end{cases}$$

Dés lors, nous avons

$$\varphi(t) = \begin{cases} \frac{\nu}{2}(t-1)^2 + \mu(t - \log t - 1) & \text{si } t > 0, \\ +\infty & \text{sinon.} \end{cases} \tag{3.7}$$

Remarque : $h(t) = \varphi_2(t)$ dans les exemples ci-dessus. Et donc, $h(t) \in \Phi_1 \cap \Phi_2$.

Voici un graphique de cette fonction φ particulière. La fonction bleue correspond au paramètre $\mu = 1$ et $\nu = 0$, et pour la fonction verte nous avons $\mu = 1$ et $\nu = 1$.

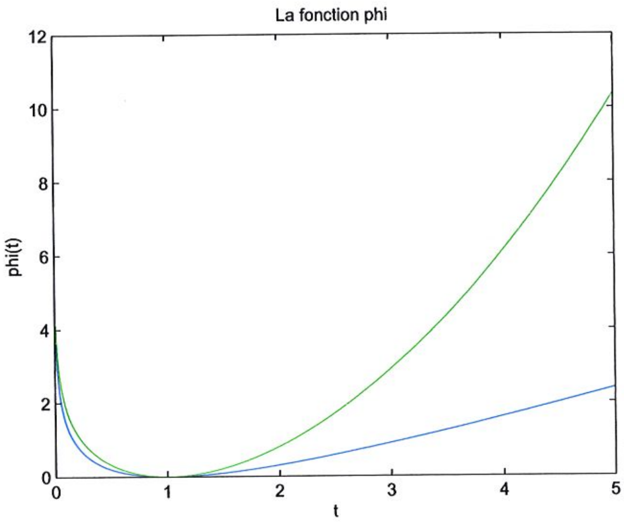

FIG. 3.1 – Fonction φ

Théorème 3.1.1 *Soit φ donnée par (3.7). Alors,*

(i) $\varphi \in \Phi_1 \cup \Phi_2$.

(ii) φ *est fortement convexe sur* $I\!R_{++}$ *avec un module* $\nu > 0$.

(iii) La fonction conjuguée de φ *est donnée par*

$$\varphi^\star(s) = \frac{\nu}{2}t^2(s) + \mu \log t(s) - \frac{\nu}{2}, \tag{3.8}$$

$$t(s) := (2\nu)^{-1}\{(\nu-\mu)+s+\sqrt{((\nu-\mu)+s)^2+4\nu\mu}\} = (\varphi^\star)'(s). \tag{3.9}$$

(iv) dom $\varphi^\star = I\!R$, *et* $\varphi^\star \in C^\infty(I\!R)$.

(v) $(\varphi^\star)'(s) = (\varphi')^{-1}(s)$ *est Lipschitz pour tout* $s \in I\!R$, *avec la constante* ν^{-1}.

(vi) $(\varphi^\star)''(s) \leq \nu^{-1}, \forall s \in I\!R$.

Preuve :

(i) et (ii) sont immédiats par la définition de φ en (3.7).

(iii) La fonction conjuguée de φ est définie par $\varphi^\star(s) = \sup_{t>0}\{ts - \varphi(t)\}$, et le supremum est atteint lorsque

$$s = \varphi'(t), \quad t > 0. \tag{3.10}$$

Utilisant la définition de φ, la relation (3.10) se réduit à

$$\begin{aligned} \varphi'(t) &= s \\ \nu(t-1) + \mu(1-\frac{1}{t}) &= s \\ \nu t^2 - t(\nu - \mu + s) - \mu &= 0, \end{aligned} \tag{3.11}$$

qui admet l'unique racine carrée positive

$$t(s) := (2\nu)^{-1}\{(\nu-\mu)+s+\sqrt{((\nu-\mu)+s)^2+4\nu\mu}\}. \tag{3.12}$$

Puisque $(\varphi^\star)' = (\varphi')^{-1}$, nous avons donc aussi de (3.10), $t(s) = (\varphi^\star)'(s)$, et

$$\begin{aligned} (\varphi^\star)(s) =& t(s)s - \varphi(t(s)) \\ =& \nu t(s)^2 - \nu t(s) + \mu t(s) - \mu - \varphi(t(s)) \text{ de (3.11)} \\ =& \frac{\nu}{2}t^2(s) + \mu \log t(s) - \frac{\nu}{2}, \end{aligned}$$

ce qui prouve (iii) et par conséquent (iv).

(v) Puisque par (ii) φ est fortement convexe de module ν, alors,

$$(t_1 - t_2)(\varphi'(t_1) - \varphi'(t_2)) \geq \nu(t_1 - t_2)^2 \quad \forall t_1,\, t_2 > 0.$$

Posons $\varphi'(t_1) = s_1$, $\varphi'(t_2) = s_2$, et comme nous avons que $(\varphi^\star)' = (\varphi')^{-1}$ l'inégalité ci-dessus se réduit à

$$((\varphi^\star)'(s_1) - (\varphi^\star)'(s_2))(s_1 - s_2) \geq \nu((\varphi^\star)'(s_1) - (\varphi^\star)'(s_2))^2,$$

et d'où $|(\varphi^\star)'(s_1) - (\varphi^\star)'(s_2)| \leq \nu^{-1}|s_1 - s_2|$, $\forall s_1,\, s_2 \in \mathbb{R}$, prouvant (v) et alors (vi) suit immédiatement. □

3.2 L'algorithme

Nous considérons le problème convexe non différentiable à contrainte linéaire

$$(P) \qquad f_\star = \inf\{f(x) : x \in C\},$$

où $f : \mathbb{R}^p \to \mathbb{R} \cap \{+\infty\}$ est une fonction propre convexe s.c.i et C est un ensemble polyédral non vide donné par

$$C = \{x \in \mathbb{R}^p : Ax \leq b\},$$

avec A une matrice $m \times p \quad (m \geq p)$ de rang maximal et $b \in \mathbb{R}^m$. Nous désignons a_i les lignes de la matrice A.
Associée à φ nous définissons la fonction distance

$$d_\varphi(u, v) = \sum_{j=1}^{p} v_j^2 \varphi(v_j^{-1} u_j), \quad \forall u,\, v \in \mathbb{R}^p_{++}, \tag{3.13}$$

Cette distance d_φ a les propriétés basiques suivantes :

(i) d_φ est une fonction homogène d'ordre 2, c'est-à-dire,

$$d_\varphi(\alpha x, \alpha y) = \alpha^2 d_\varphi(x, y), \quad \forall \alpha > 0.$$

(ii) $\forall (x,\, y) \in \mathbb{R}^p_{++} \times \mathbb{R}^p_{++}$ nous avons :

$$d_\varphi \geq 0 \quad \text{et,}$$
$$d_\varphi(x, y) = 0 \text{ si } x = y.$$

La première propriété est évidente par la définition (3.13), et la seconde suit du fait que $\varphi(t) \geq 0$ et $\varphi(t) = 0$ si et seulement si $t = 1$.

De plus pour $x,\, y \in \operatorname{int} C$, notons

$$l_i(x) = b_i - \langle a_i,\, x\rangle, \quad i = 1, \ldots, p, \tag{3.14}$$

$$L(x) = (l_1(x), \dots, l_m(x)), \tag{3.15}$$
$$d_1(x,y) = \nu d_\varphi(L(x), L(y)), \tag{3.16}$$
$$d_2(x,y) = \frac{\nu}{2}\|L(x) - L(y)\|^2 = \frac{\nu}{2}\|A(x-y)\|^2, \tag{3.17}$$
$$d(x,y) = d_1(x,y) + d_2(x,y). \tag{3.18}$$

Remarquons que parce que A est de rang maximal, $(u, v) \longrightarrow \langle A^T Au, v\rangle$ définit un produit scalaire sur $\mathbb{R}^p$ désigné par $(u, v)_A$ avec $\|u\|_A := \|Au\| = \langle Au, Au\rangle^{1/2}$, donc nous pouvons écrire

$$d(x,y) = d_1(x,y) + \frac{\nu}{2}\|x-y\|_A^2. \tag{3.19}$$

La Proposition suivante collecte des propriétés intéressantes de d défini précédemment.

Proposition 3.2.1 *Supposons d défini en (3.18). Alors pour tout $x, y \in$ int C :*

(a) $d(x,y) \geq 0$ et égale 0 si et seulement si $x = y$;

(b) $x \longrightarrow d(x,y)$ est fortement convexe de module ν, et donc

$$\langle \nabla_x d(x,p) - \nabla_x d(y,p),\, x-y\rangle \geq \nu\|x-y\|_A^2, \quad \forall p \in \; int\, C;$$

(c)

$$\langle \nabla_x d(x,y),\, z-x\rangle \leq \theta(\|z-y\|_A^2 - \|z-x\|_A^2), \quad \forall z \in C; \tag{3.20}$$

(d) $\|x-y\|_A^2 \geq \lambda_{min}(A^TA)\|x-y\|^2$, où $\lambda_{min}(A^TA)$ est la plus petite valeur propre de la matrice A^TA symétrique définie positive.

Preuve :

(a) et (b) suivent de la convexité de φ, la définition de d et de l'hypothèse de rang plein sur A.
(c) Nous utilisons le Lemme 3.1.2. Car nous avons bien que $h \in \Phi_1 \cap \Phi_2$ et comme $h"(1) = 1$ nous avons que $\nu \geq \mu h"(1) > 0$ et $\theta = \dfrac{v + \mu h"(1)}{2}$
(d) suit de la norme induite sur par $\|\cdot\|_A$ et de l'hypothèse de rang plein sur A. □

Proposition 3.2.2 *Soit $g \in \mathbb{R}^p$, et soit $x \in$ int C. Alors pour chaque $\lambda > 0$, il existe un point $x(\lambda) \in$ int C unique satisfaisant*

$$x(\lambda) := arg\ min\ _u\{\lambda\langle g,\, u\rangle + d(u,x)\}.$$

Preuve :

Soit $x \in$ int C. L'existence et l'unicité de $x(\lambda)$ est garantie par la forte convexité de d établit dans la Proposition précédente (b), et puisque dom $d(\cdot, x) =$ int C, il suit que $x(\lambda) \in$ int C. $\square$

Lemme 3.2.1 *Soit $\{v_k\}$ et $\{\beta_k\}$ des suites de nombres réels positifs satisfaisant*

(i) $v_{k+1} \leq v_k + \beta_k$,

(ii) $\sum_{k=1}^{\infty} \beta_k < \infty$.

Alors, la suite $\{v_k\}$ converge.

Lemme 3.2.2 *Soit $\{\lambda_k\}$ une suite de nombres positifs, $\{a_k\}$ une suite de nombres réels convergeant vers un certain a, et $b_n := \sigma_n^{-1} \sum_{k=1}^{n} \lambda_k a_k$, où $\sigma_n = \sum_{k=1}^{n} \lambda_k$. Si $\sigma_n \to \infty$, alors $b_n \to a$.*

Voici maintenant l'algorithme.

La méthode proximale logarithmique-quadratique (LQP).

Algorithme 3.2.1 *Prendre comme point de départ le point $x^0 \in$ int C, $\epsilon_k > 0$, $\lambda_k > 0$ et on génère une suite $\{x^k\} \in$ int C, satisfaisant*

$$g^k \in \partial_{\epsilon_k} f(x^k), \tag{3.21}$$

$$\lambda_k g^k + \nabla_x d(x^k, x^{k-1}) = 0, \tag{3.22}$$

où $\partial_{\epsilon_k} f(x^k)$ est le ϵ_k-sous-différentiel de f en x^k. Nous notons $X_\star$ l'ensemble des solutions optimales du Problème (P).

Trois relations importantes pour notre analyse peuvent être déduites de cet algorithme.

$$\lambda_k(f(x^k) - f(x)) \leq \langle x - x^k,\, \nabla_x d(x^k,\, x^{k-1})\rangle + \lambda_k \epsilon_k, \tag{3.23}$$

$$f(x) + \lambda_k^{-1} d(x,\, x^{k-1}) \geq f(x^k) + \lambda_k^{-1} d(x^k,\, x^{k-1}) - \epsilon_k \quad \forall x, \tag{3.24}$$

$$f(x^{k-1}) - f(x^k) \geq \lambda_k^{-1} d(x^k,\, x^{k-1}) - \epsilon_k. \tag{3.25}$$

La première inégalité suit de (3.21) et (3.22) et de l'inégalité du ϵ−sous-gradient pour une fonction convexe, la seconde inégalité suit de la première avec la définition de d et la convexité de $d(\cdot, x^{k-1})$, et la dernière est un cas spécial de la précédente avec $x = x^{k-1}$, puisque $d(x^{k-1}, x^{k-1}) = 0$.

Théorème 3.2.1 *Supposons que dom $f \cap$ int $C \neq \emptyset$. Soit $\{x^k\}$ une suite générée par l'algorithme LQP. Soit aussi $\{\lambda_k\}$ une suite arbitraire de nombres positifs et posons $\sigma_n := \sum_{i=1}^{n} \lambda_k$. Alors,*

(i) $\lambda_k(f(x^k) - f(x)) \leq \theta[\|(x - x^{k-1})\|_A^2 - \|(x - x^k)\|_A^2] + \lambda_k \epsilon_k, \quad \forall x \in C;$

(ii) $f(x^n) - f(x) \leq \theta \sigma_n^{-1} \|x - x^0\|^2 + \sigma_n^{-1} \sum_{k=1}^{n} \sigma_k \epsilon_k, \quad \forall x \in C;$

(iii) Si $\sigma_n \to \infty$ et $\epsilon_k \to 0$, alors $\liminf_{n\to\infty} f(x^n) \to f_\star$. De plus, si $\sum_{k=1}^{\infty} \epsilon_k < \infty$ alors, $f(x^n) \to f_\star = \inf\{f(x) : x \in C\}$ si $f_\star$ est finie.

(iv) Si l'ensemble optimal $X_\star \neq \emptyset$, $\sigma_n \to \infty$, et $\sum_{k=1}^{n} \lambda_k \epsilon_k < \infty$, alors la suite $\{x^n\}$ converge vers la solution optimale de (P).
Dans le cas spécial où $C = I\!R_+^p$, si $\lambda_k \geq \lambda > 0$, alors

$$\liminf_{k\to\infty} g_i^k \geq 0, \qquad \lim_{k\to\infty} g_i^k x_i^k = 0, \quad \forall i. \tag{3.26}$$

Preuve :

(i) En utilisant (3.23), nous avons que

$$\lambda_k(f(x^k) - f(x)) \leq \langle x - x^k, \nabla_x d(x^{k-1}, x^k)\rangle + \lambda_k \epsilon_k.$$

Et donc il suffit de prouver que

$$\langle x - x^k, \nabla_x d(x^k, x^{k-1})\rangle \leq \theta[\|x - x^{k-1}\|_A^2 - \|x - x^k\|_A^2].$$

En appliquant la Proposition 3.2.1 (c) nous obtenons

$$\lambda_k(f(x^k) - f(x)) \leq \theta(\|x - x^{k-1}\|^2 - \|x - x^k\|^2) + \lambda_k \epsilon_k. \tag{3.27}$$

(ii) En sommant (3.27) sur les $k = 1, \ldots, n$ nous obtenons

$$-\sigma_n f(x) + \sum_{k=1}^{n} \lambda_k f(x^k) \leq \theta[\|x - x^0\|^2 - \|x - x^n\|^2] + \sum_{k=1}^{n} \lambda_k \epsilon_k. \tag{3.28}$$

Maintenant, de la relation (3.25) nous avons (puisque $d_\varphi \geq 0$),

$$f(x^k) - f(x^{k-1}) \leq \epsilon_k. \tag{3.29}$$

Puisque $\sigma_n = \sum_{k=1}^n \lambda_k$, utilisant $\sigma_k = \lambda_k + \sigma_{k-1}$ (avec $\sigma_0 \equiv 0$), multipliant l'inégalité ci-dessus par σ_{k-1} et sommant sur $k = 1, \ldots, n$ nous obtenons

$$\sum_{k=1}^{n}(\sigma_k - \lambda_k)f(x^k) - \sigma_{k-1}f(x^{k-1}) \leq \sum_{k=1}^{n}\sigma_{k-1}\epsilon_k,$$

ce qui se réduit à

$$\sigma_n f(x^n) - \sum_{k=1}^{n}\lambda_k f(x^k) \leq \sum_{k=1}^{n}\sigma_{k-1}\epsilon_k. \tag{3.30}$$

Additionnant (3.28) et (3.30) nous obtenons

$$\sigma_n(f(x^n) - f(x)) \leq \theta[\|x - x^0\|^2 - \|x - x^n\|^2] + \sum_{k=1}^{n}\sigma_k\epsilon_k,$$

puisque $\lambda_k + \sigma_{k-1} = \sigma_k$ et donc (ii) suit.

(iii) De (3.28) nous obtenons

$$\sigma_n^{-1}\sum_{k=1}^{n}\lambda_k f(x^k) \leq f(x) + \theta\sigma_n^{-1}\|x - x^0\|^2 + \sigma_n^{-1}\sum_{k=1}^{n}\lambda_k\epsilon_k. \tag{3.31}$$

Passant à la limite lorsque $n \to +\infty$, puisque $\sigma_n \to +\infty$, invoquant le Lemme 3.2.2 avec $\epsilon_k \to 0$ il suit de (3.31) que

$$\lim\inf_{n\to+\infty} f(x^n) \leq \inf\{f(x) : x \in int\, C\},$$

ce qui implique avec $f(x^n) \geq \inf\{f(x) : x \in C\} = f_\star$ que

$$\lim\inf_{n\to+\infty} f(x^n) = \inf\{f(x) : x \in C\} = f_\star.$$

Maintenant, de (3.29), quand $f_\star$ est finie, nous avons aussi que

$$(f(x^k) - f_\star) - (f(x^{k-1}) - f_\star) \leq \epsilon_k.$$

$$f(x^k) - f_\star \leq \epsilon_k + (f(x^{k-1}) - f_\star)$$

Puisque $f(x^k) \geq f_\star$, $\forall k$, et $\sum \epsilon_k < \infty$, invoquant le Lemme 3.2.1 nous avons que la $\lim_{n\to\infty} f(x^n)$ existe et (iii) est prouvé.

(iv) Soit $x^\star \in X_\star \neq \emptyset$. Puisque

$$f(x^k) \geq f(x^\star) \quad \forall k, \tag{3.32}$$

nous obtenons de (3.27),

$$\|x^\star - x^k\|^2 \leq \|x^\star - x^{k-1}\|^2 + \sigma^{-1}\lambda_k\epsilon_k.$$

Puisque $\sum_{k=1}^{\infty} \lambda_k \epsilon_k < \infty$, alors par le Lemme 3.2.1, $\{\|x^\star - x^k\|\}$ converge, et la suite $\{x^k\}$ est bornée. Soit $\{x^{k_j}\}$ une sous-suite convergeant vers $x^\infty \in C$. Comme $f(x^k) \to f_\star$, nous avons que $f(x^{k_j}) \to f_\star$ et d'où avec f restant fermé nous avons $f(x^\infty) \leq \lim_{k_j \to \infty} f(x^{k_j}) = f_\star$ et il suit que $x^\infty \in X_\star$. Donc, $\|x^\infty - x^k\|$ converge. Puisque $\{x^{k_j}\} \in \operatorname{int} C$ converge vers $x^\infty \in C$, $\|x^\infty - x^{k_j}\| \to 0$ et d'où $\|x^\infty - x^k\| \to 0$, ce qui implique grâce aux propriétés de la norme que le point limite est unique et $x^k \to x^\infty \in X_\star$.

Pour prouver la dernière partie de (iv), nous sommes dans le cas particulier où $C = \mathbb{R}_+^p$, notons que (3.22) est équivalent à

$$g_i^k = \lambda_k^{-1}(\nu(x_i^{k-1} - x_i^k) - \mu x_i^{k-1} h'(x_i^k / x_i^{k-1})), \quad i = 1, \ldots, p. \tag{3.33}$$

Puisque $h \in \Phi_2$, de (3.2),

$$-\mu x_i^{k-1} h'(x_i^k / x_i^{k-1}) \geq \mu h"(1)(x_i^{k-1} - x_i^k), \quad i = 1, \ldots, p$$

et alors il suit de (3.33) que

$$g_i^k \geq \lambda_k^{-1}(\mu h"(1) + \nu)(x_i^{k-1} - x_i^k) = \frac{\theta}{2} \lambda_k^{-1}(x_i^{k-1} - x_i^k), i = 1, \ldots, p. \tag{3.34}$$

Puisque $\{\lambda_k^{-1}\}$ est bornée supérieurement et que $\{x^k\}$ converge, il suit que

$$\lim_{k \to \infty} \inf g_i^k \geq 0, \quad \forall i = 1, \ldots, p$$

De plus, utilisant le coté gauche de l'inégalité (3.2), nous avons

$$-\mu x_i^{k-1} h'(x_i^k / x_i^{k-1}) \leq \mu h"(1)(x_i^{k-1} - x_i^k) \frac{x_i^{k-1}}{x_i^k}, \quad i = 1, \ldots, p.$$

Utilisant (3.34) il suit que

$$\lambda_k^{-1} x_i^k (x_i^{k-1} - x_i^k)(\mu h"(1) + \nu) \leq g_i^k x_i^k$$

$$\leq \lambda_k^{-1}(x_i^{k-1} - x_i^k)(\mu h"(1) x_i^{k-1} + \nu x_i^k), \quad i = 1, \ldots, p.$$

Donc, une fois encore puisque $\{\lambda_k^{-1}\}$ est bornée supérieurement et que $\{x^k\}$ converge il suit que

$$\lim_{k \to \infty} g_i^k x_i^k = 0 \quad \forall i = 1, \ldots, p.$$

□

Chapitre 4

Méthode de faisceaux intérieure.

4.1 Préliminaires.

Soit ϵ−argmin $F(x) := \{z : F(x) \leq \inf F + \epsilon\}$, avec F une fonction donnée et $\epsilon \geq 0$, définissons

$$F_k(x) = f(x) + \lambda_k^{-1} d_\varphi(x, x^{k+1}). \tag{4.1}$$

Lemme 4.1.1 *Soit $x^{k-1} \in I\!\!R^p_{++}$, $\lambda_k > 0$ et F_k défini en (4.1). Si une des hypothèses suivantes est satisfaite :*

(i) L'ensemble optimal $X_\star$ est compact et non vide,

(ii) φ est cofinie, c'est-à-dire $\varphi_\infty(d) = +\infty$, $\forall d \neq 0$.

Il existe une solution unique

$$y^k = \mathit{argmin}\ _x F_k(x), \tag{4.2}$$

$$\lambda_k g^k + \Phi'(x^{k-1}, y^k) = 0, \quad g^k \in \partial f(y^k). \tag{4.3}$$

où $\Phi'(a,b) := (a_1\varphi'(b_1/a_1), \ldots, a_p\varphi'(b_p/a_p))^T \quad \forall a, b \in I\!\!R^p_{++}$

Preuve :

Comme la fonction F_k définie en (4.1) est fortement convexe, nous avons l'unicité de l'ensemble optimal $S_k := \text{argmin}_x F_k$. Si S_k est non vide, puisque $\varphi \in \Phi$, nous avons que $\lim_{t \to 0^+} \varphi'(t) = -\infty$ et avec l'hypothèse que dom $f \cap$ $\mathbb{R}^p_{++} \neq 0$ il suit que $y^k \in \mathbb{R}^p_{++}$, et alors puisque $0 \in \partial F_k(y_k)$, nous obtenons (4.3).
Il reste à prouver que avec (i) ou (ii), l'ensemble optimal S_k est non vide

et compact. Utilisant [8] (Théorème 27.1 (d)), nous avons que S_k est un ensemble non vide compact si

$$(F_k)_\infty(d) > 0, \quad \forall d \neq 0. \tag{4.4}$$

De (4.1) nous avons :

$$(F_k)_\infty(d) = f_\infty(d) + \delta(d|\mathbb{R}^p_+) + d_\varphi(\cdot, x^{k-1})_\infty(d), \tag{4.5}$$

où $\delta(d|\mathbb{R}^p_+)$ est la fonction indicatrice de $\mathbb{R}^p_+$.
Si nous avons l'hypothèse (i), c'est-à-dire $X_\star$ est un ensemble non vide compact, alors par [8] (Théorème 27.1 (d)), nous avons que $f_\infty + \delta(d|\mathbb{R}^p_+) > 0, \forall d \neq 0$. De plus puisque

$$d_\varphi(\cdot, x^{k-1})_\infty(d) = \lim_{t \to +\infty} t^{-1} d_\varphi(x^{k-1} + td, x^{k-1}) \geq 0,$$

il suit de (4.5) que (4.4) est satisfait.
Maintenant, si nous avons l'hypothèse (ii) avec φ confinie, nous obtenons de (4.5) que $(F_k)_\infty(d) = +\infty, \forall d \neq 0$, et donc (4.4) est aussi satisfait. □

4.2 Cas général.

Nous allons développer une méthode de faisceaux intérieure pour résoudre le problème convexe non différentiable à contrainte linéaire

$$\text{(P)} \quad f_\star = \inf\{f(x) : x \in C\}.$$

Et nous introduisons les hypothèses suivantes :

Hypothèses A :
(a) $C \subset \text{dom} f$.
(b) A est de rang maximal (ce qui est clairement satisfait quand $C = \mathbb{R}^p_+$).
(c) $\text{int} C \neq \emptyset$.
(d) Le sous-différentiel ∂f est borné sur les sous-ensembles bornés de $\text{int} C$.

Notons que l'hypothèse (d) est automatiquement satisfaite quand $C \subset \text{int dom } f$. Ce dernier a toujours lieu quand f est à valeurs réelles, une hypothèse qui est communément utilisée pour la plupart des méthodes de faisceaux.

Pour minimiser f sur C, nous devons d'abord résoudre le problème approximatif. Soit $\lambda > 0$, $\bar{x} \in \text{ int } C$, et considérer le problème de minimisation :

$$P_\lambda(\bar{x}) \qquad m_\lambda(\bar{x}) = \inf\{f(x) + \lambda^{-1} d(x, \bar{x}) : x \in \mathbb{R}^p\}.$$

Comme une conséquence du Lemme 4.1.1 (car nous avons que φ est cofinie), le problème $P_\lambda(\bar{x})$ admet une solution unique notée $\text{prox}^d_{\lambda f}(\bar{x}) \in C$. Maintenant nous allons définir un algorithme pour résoudre le problème $P_\lambda(\bar{x})$.

On commence avec $x_{-1} \in \ \text{int}\, C, \quad g_{-1} \in \partial f(x_{-1})$, et $\psi^0(x) = \langle g_{-1}, x - x_{-1}\rangle + f(x_{-1})$.

Algorithme 4.2.1 *A partir de ψ^0, on génère la suite $\{x_k, \psi^k\}_{k\in I\!N}$ comme suit :*
Supposons qu'à l'itération k, ψ^k est connu. Alors on calcule
Pas 1 :

- $x_k = prox^d_{\lambda\psi^k}(\bar{x})$;
- *poser* $d^k = -\lambda^{-1}\nabla_x d(x_k, \bar{x})$;
- $l^k(x) = \psi^k(x_k) + \langle d^k, x - x_k\rangle$.

Pas 2 : Soit $\psi^{k+1} : I\!R^p \to I\!R \cup +\infty$, est une fonction propre convexe s.c.i satisfaisant $\forall k \in I\!N$:
(H1) $\psi^{k+1} \leq f$,
(H2) $l^k \leq \psi^{k+1}$,
(H3) $f(x_k) + \langle g_k, x - x_k\rangle \leq \psi^{k+1}(x), \quad \forall x$ *avec* $g_k \in \partial f(x_k)$.

Nous pouvons voir en utilisant la condition d'optimalité pour $x_k = \text{prox}^d_{\lambda\psi^k}(\bar{x})$ (le "Pas 1" de l'algorithme) que

$0 \in \partial\psi^k(x_k) + \lambda^{-1}\nabla d(\bar{x}, x_k)$ et nous savons aussi que $\lambda^{-1}\nabla d(\bar{x}, x_k) = -d^k$

et donc nous obtenons

$$d^k \in \partial\psi^k(x_k). \tag{4.6}$$

Voici deux exemples de fonction ψ^k satisfaisant les hypothèses (H1) à (H3) définies ci-dessus.

$$\psi^{k+1}(x) = \max\{f(x_i) + \langle g_i, x - x_i\rangle : i \leq k\} \text{ avec } g_i \in \partial f(x_i), \tag{4.7}$$

$$\psi^{k+1} = \max\{l^k(x), f(x_k) + \langle g_k, x - x_k\rangle\} \text{ avec } g_k \in \partial f(x_k) \tag{4.8}$$

Dans le premier exemple nous remarquons que

$$\psi^{k+1}(x_j) = f(x_j), \quad \forall j \leq k. \tag{4.9}$$

Théorème 4.2.1 *Supposons avoir les hypothèses A et une suite $\{x_k, \psi^k\}$ générée par l'algorithme 4.2.1. Posons aussi $\epsilon_k := f(x_k) - \psi^k(x_k)$. Alors nous avons que*

(a) $\epsilon_k \geq 0$, *et* $\lim_{k\to\infty} \epsilon_k = 0$.

(b) $d^k \in \partial_{\epsilon_k} f(x_k)$, *le* ϵ_k*-sous-différentiel de* f *en* x_k.

(c) La suite $\{x_k\}$ *converge vers* $prox^d_{\lambda f}(\bar{x})$.

Preuve :

(a) Pour tout k et tout $x \in \text{int } C$, nous définissons

$$\hat{l}^k(x) := l^k(x) + \lambda^{-1} d(x, \bar{x}),$$

$$\hat{\psi}^k(x) := \psi^k(x) + \lambda^{-1} d(x, \bar{x}).$$

Parce que $d(x, \bar{x}) \geq 0$, de H1, de la définition de $\hat{\psi}^{k+1}$ et de la définition de x_{k+1}, nous avons que

$$f(\bar{x}) \geq \psi^{k+1}(\bar{x}) = \hat{\psi}^{k+1}(\bar{x}) \geq \hat{\psi}^{k+1}(x_{k+1}), \tag{4.10}$$

et de la définition de l^k et l'hypothèse H2, nous obtenons

$$\hat{\psi}^{k+1}(x_{k+1}) = \hat{l}^{k+1}(x_{k+1}) \geq \hat{l}^k(x_{k+1}). \tag{4.11}$$

De plus, nous avons que

$$\begin{aligned}
\hat{l}^k(x) - \hat{l}^k(x_k) &= l^k(x) + \lambda^{-1} d(x, \bar{x}) - l^k(x_k) - \lambda^{-1} d(x_k, \bar{x}) \\
&= l^k(x) - l^k(x_k) + \lambda^{-1}(d(x, \bar{x}) - d(x_k, \bar{x})) \\
&= \psi^k(x_k) + \langle d^k, x - x_k \rangle - \psi^k(x_k) - \langle d^k, x_k - x_k \rangle + \lambda^{-1}(d(x, \bar{x}) - d(x_k, \bar{x})) \\
&= \langle d^k, x - x_k \rangle + \lambda^{-1}(d(x, \bar{x}) - d(x_k, \bar{x})).
\end{aligned}$$

Utilisant le fait que $d^k = -\nabla_x d(x_k, \bar{x})$ et définissant

$$a = \lambda^{-1}\mu(\langle -\nabla_x d_1(x_k, \bar{x}), x - x_k \rangle + d_1(x, \bar{x}) - d_1(x_k, \bar{x})),$$

$$b = \lambda^{-1}\nu(\langle \bar{x} - x_k, x - x_k \rangle_A + 2^{-1}(\|x - \bar{x}\|_A^2 - \|x_k - \bar{x}\|_A^2)),$$

nous pouvons écrire

$$\langle d^k, x - x_k \rangle + \lambda^{-1}(d(x, \bar{x}) - d(x_k, \bar{x})) = a + b.$$

Clairement, nous avons que $b = (2\lambda)^{-1}\nu\|x - x_k\|_A^2$, et parce que $d_1(\cdot, \bar{x})$ est convexe, il suit que $a \geq 0$, et nous obtenons

$$\hat{l}^k(x) - \hat{l}^k(x_k) \geq (2\lambda)^{-1}\nu\|x - x_k\|_A^2, \tag{4.12}$$

et donc

$$\hat{l}^k(x_{k+1}) - \hat{l}^k(x_k) \geq (2\lambda)^{-1}\nu\|x_k - x_{k+1}\|_A^2 \geq 0, \tag{4.13}$$

qui combiné avec (4.10)- (4.11) donne

$$f(\bar{x}) \geq \hat{l}^{k+1}(x_{k+1}) \geq \hat{l}^k(x_k).$$

Par conséquent, la suite $\{\hat{l}^k(x_k)\}$ converge vers un $l^\star \in \mathbb{R}$, et de (4.13) nous obtenons

$$\lim_{k\to\infty} \|x_k - x_{k+1}\|_A = 0. \tag{4.14}$$

De plus, de H1 et H3, nous avons

$$f(x_{k+1}) - f(x_k) \geq \psi^{k+1}(x_{k+1}) - f(x_k) \geq \langle g_k,\, x_{k+1} - x\rangle. \tag{4.15}$$

parce que de H1 et H2 nous avons $l^k(\bar{x}) \leq f(\bar{x})$, il suit que $\hat{l}^k(\bar{x}) \leq f(\bar{x})$. Alors, utilisant (4.12) avec $x = \bar{x}$ et le fait que $\{\hat{l}^k(x_k)\}$ converge vers $l^\star$, il suit que la suite $\{x_k\}$ est bornée. Maintenant, parce que $x_k \in \operatorname{int} C \subset \operatorname{int} \operatorname{dom} f$, utilisant le Théorème des accroissements finis, il existe y_k sur le segment de droite ouvert $]x_k,\, x_{k+1}[$ avec $c_k \in \partial f(y_k)$ tel que $f(x_{k+1}) - f(x_k) = \langle c_k,\, x_{k+1} - x_k\rangle$. Alors, parce que ∂f est bornée sur les sous- ensembles de int C, il suit que les suites $\{c_k\}$ et $\{g_k\}$ sont bornées, et avec (4.15) nous obtenons

$$\lim_{k\to\infty} (f(x_{k+1}) - f(x_k)) = \lim_{k\to\infty} (\psi^{k+1}(x_{k+1}) - f(x_k)) = 0. \tag{4.16}$$

Maintenant par H1, nous avons

$$0 \leq \underbrace{f(x_k) - \psi^k(x_k)}_{=\epsilon_k} = \underbrace{f(x_k) - f(x_{k-1})}_{\to 0} + \underbrace{f(x_{k-1}) - \psi^k(x_k)}_{\to 0},$$

et donc la partie (a) du Théorème suit de (4.16).

(b) Parce que ψ^k est convexe, et utilisant H1 il suit de (4.6) que

$$\begin{aligned}\langle d^k,\, x - x_k\rangle \leq \psi^k(x) - \psi^k(x_k) &\leq f(x) - \psi^k(x_k)\\ &\leq f(x) - f(x_k) + f(x_k) - \psi^k(x_k)\\ &= f(x) - f(x_k) + \epsilon_k.\end{aligned}$$

Comme une conséquence de la dernière relation et de la définition du ϵ_k-sous-différentiel, nous avons que $d^k \in \partial_{\epsilon_k} f(x_k)$ et (b) est prouvé.

(c) Nous devons prouver que $\{x_k\}$ converge vers $\operatorname{prox}^d_{\lambda f}(\bar{x})$. Par (b) nous avons que

$$0 \in \partial_{\epsilon_k} f(x_k) + \lambda^{-1}\nabla_x d(x_k, \bar{x}),$$

et par définition de $\operatorname{prox}^d_{\lambda f}(\bar{x})$ nous pouvons écrire

$$0 \in \partial_{\epsilon_k} f(\operatorname{prox}^d_{\lambda f}(\bar{x})) + \lambda^{-1}\nabla_x d(\operatorname{prox}^d_{\lambda f}(\bar{x}), \bar{x}).$$

Des deux inclusions ci-dessus et de la définition du ϵ-sous-gradient d'une fonction convexe f, il suit que

$$\langle \nabla_x d(\operatorname{prox}^d_{\lambda f}(\bar{x}), \bar{x}) - \nabla_x d(x_k, \bar{x}), \operatorname{prox}^d_{\lambda f}(\bar{x}) - x_k \rangle \leq 2\lambda\epsilon_k.$$

Invoquant la Proposition 3.2.1 (b), nous obtenons

$$\nu \| \operatorname{prox}^d_{\lambda f}(\bar{x}) - x_k \|_A^2 \leq 2\lambda\epsilon_k,$$

et il suit alors que lorsque $\epsilon_k \to 0$, la suite $\{x_k\}$ converge vers $\operatorname{prox}^d_{\lambda f}(\bar{x})$. $\square$

Nous allons maintenant présenter un algorithme pour résoudre le problème initial (P) :

$$\text{(P)} \qquad f_\star = \inf\{f(x) : x \in C\}.$$

Soit $\eta \in (0,1)$ un paramètre fixé.

Algorithme de faisceaux intérieur (IBA).

Algorithme 4.2.2 *Soit $\{\lambda_k\}$ une suite de nombres positifs. On commence avec $y^0 \in int\ C$ et on génère la suite $\{y^k\}, k = 0, 1, \ldots,$ comme suit :*
Pas 1 : Si y^k est optimale pour (P), alors on arrête. Sinon :
Pas 2 : Calculer via l'algorithme 4.2.1, commençant avec le point y^k et en un nombre de pas $n(k)$, un point $x_{n(k)}$, satisfaisant

$$x_{n(k)} = prox^d_{\lambda_k \psi^{n(k)}}(y^k), \tag{4.17}$$

$$f(y^k) - f(x_{n(k)}) \geq \eta(f(y^k) - \psi^{n(k)}(x_{n(k)})). \tag{4.18}$$

Pas 3 : Poser $y^{k+1} = x_{n(k)}$ et aller au Pas 1.

Théorème 4.2.2 *Supposons que le Problème (P) admette une solution optimale et que les hypothèses A sont satisfaites. Soit $\{y^k\}$ la suite générée par l'algoritme IBA. Alors*

(a) L'algorithme IBA est bien défini ; c'est-à-dire, si y^k n'est pas une solution optimale, alors l'algorithme IBA donne en un nombre fini de pas $n(k)$ un point $x_{n(k)}$, satisfaisant la relation (4.17). De plus, nous avons

$$d^k := -\lambda_k^{-1}\nabla_x d(y^{k+1}, y^k), \qquad d^k \in \partial\psi^{n(k)}(y^{k+1}).$$

(b) Si $0 < \lambda \leq \lambda_k \leq \bar{\lambda}$, alors $y^k \to y^\star \in arg\ min_C f$.

Preuve :

(a) Par l'absurde. Nous supposons que y^n n'est pas une solution optimale de (P) et que le Pas 1 échoue à l'étape n. Alors, pour chaque k, nous avons

$$f(y^n) - f(x_k) < \eta(f(y^n) - \psi^k(x_k)); \quad x_k = \text{prox}^d_{\lambda_n \psi^k}(y^n),$$

et donc

$$\begin{aligned} f(y^n) - f(x_k) &< \eta(f(y^n) - f(x_k) + f(x_k) - \psi^k(x_k)); \\ (1-\eta)(f(y^n) - f(x_k)) &< \eta(f(x_k) - \psi^k(x_k)) := \epsilon_k. \end{aligned} \tag{4.19}$$

Comme $z_n := \text{prox}^d_{\lambda_n f}(y^n) \in \text{int } C \subset \text{ int dom } f$, f est continue en z_n, et parce que par le Théorème 4.2.1 la suite $\{x_k\}$ converge vers z_n, il suit que $f(z_n) = \lim_{k\to\infty} f(x_k)$.
Maintenant, par définition de z_n, nous avons $f(z_n) \leq f(y^n)$. Finalement, utilisant (4.19), parce que $\epsilon_k \to 0$ par le Théorème 4.2.1, il suit que $f(y^n) \leq \liminf_{k\to\infty} f(x_k) = f(z_n)$. Et donc nous avons que $f(\text{prox}^d_{\lambda_n f}(y^n)) = f(y^n)$. Parce que le problème $P_{\lambda_n}(y^n)$ admet une solution unique, il suit que $\text{prox}^d_{\lambda_n f}(y^n) = y^n$. Donc, parce que $\nabla_x d(y^n, y^n) = 0$, les conditions nécessaires d'optimalité du problème $P_{\lambda_n}(y^n)$ implique $0 \in \partial f(y^n)$, ce qui signifie que y^n est une solution optimale de (P). Nous avons donc une contradiction.

(b) Maintenant, nous pouvons utiliser (4.18) avec $y^{k+1} = x_{n(k)}$, et nous obtenons

$$f(y^k) - f(y^{k+1}) \geq \eta(f(y^k) - f(y^{k+1})) + \eta(f(x_{n(k)}) - \psi^{n(k)}(x_{n(k)})). \tag{4.20}$$

Posons $\epsilon_k := f(x_{n(k)}) - \psi^{n(k)}(x_{n(k)})$. Alors par le Théorème précédent, $\epsilon_k \geq 0$. De plus, (4.20) peut s'écrire comme

$$f(y^k) - f(y^{k+1}) \geq \eta(1-\eta)^{-1}\epsilon_k; \tag{4.21}$$

et finalement, parce que $d^{n(k)} = -\nabla_x d(x_{n(k)}, y^k)\lambda_k^{-1}$, il suit du Théorème 4.2.1 (b) que

$$0 \in \lambda_k^{-1}\nabla_x d(y^{k+1}, y^k) + \partial_{\epsilon_k} f(y^{k+1}),$$

donc $\{y^k\}$ n'est rien d'autre que la suite générée par l'algorithme LQP 3.2.1 pour résoudre le problème à contrainte linéaire (P). Invoquant le Théorème 3.2.1, il suit que $\{y^k\}$ converge vers la solution optimale si nous pouvons prouver que $\sum \lambda_k \epsilon_k < +\infty$. Cependant, parce que nous avons l'hypothèse que $0 < \lambda_k \leq \bar{\lambda}$, nous devons seulement montrer que $\sum \epsilon_k < +\infty$. Pour cela, nous considérons (4.21). Comme $\epsilon_k \geq 0$, il suit que la suite $\{f(y^k)\}$ est décroissante, et parce qu'elle est bornée inférieurement par $f_\star$, elle converge vers un certain $l \in \mathbb{R}$. Ensuite sommant sur k en (4.21), il suit que $\sum_{k=1}^{\infty} \epsilon_k < +\infty$. □

4.3 Le cas particulier $C = \mathbb{R}^p_+$.

Dans ce cas, nous considérons une méthode de faisceaux avec le choix particulier suivant :

$$\psi^{k+1}(x) = \max\{l^k(x), f(x_k) + \langle g_k, x - x_k\rangle\}.$$

L'avantage de ce choix est que, à chaque itération, nous devons seulement calculer

$$f(x_{k-1}), \qquad g_{k-1} \in \partial f(x_{k-1}), \qquad x_k = \mathrm{prox}^d_{\lambda\psi^k}(\bar{x}).$$

Le calcul de $f(x_{k-1})$, g_{k-1} dépend de la nature du problème.
Dans le cas d'un problème dual convexe séparable à n variables, il consiste à résoudre, à chaque itération, n sous problèmes qui minimisent une fonction d'une seule variable. Ce cas est typique dans toutes les méthodes de décomposition pour les problèmes convexes séparables.
Notre méthode sera différente des autres techniques de faisceaux si le calcul de $x_k := \mathrm{prox}^d_{\lambda\psi^k}(\bar{x})$ peut être fait efficacement. Ici, le choix de la distance d et le fait que ψ^k est le supremum de deux fonctions affines est crucial. Pour simplifier, nous supposons que $\lambda = 1$, et nous notons

$$\psi^k(x) = \max\{\langle a^1_k, x\rangle - b^1_k, \langle a^2_k, x\rangle - b^2_k\}.$$

Nous devons alors résoudre le problème

$$(\mathrm{P}_k) \qquad m_k(\bar{x}) = \inf\{\psi^k(x) + d(x, \bar{x}) \;:\; x \in \mathbb{R}^p_+\}$$

avec la solution optimale x_k. Pour un $\bar{x} \in \mathbb{R}^p_{++}$ fixé, nous posons $\Phi(x) = d(x, \bar{x})$. Nous pouvons associer au problème (P_k) son dual de Fenchel donné par

$$(\mathrm{D}_k) \qquad p_k(\bar{x}) = \inf\{(\psi^k)^\star(u) + \Phi^\star(-u) \;:\; u \in \mathbb{R}^p\}, \tag{4.22}$$

où $\Phi^\star$ est la conjuguée de Φ.

Le résultat suivant prouve que (D_k) peut être simplifié et réduit à la minimisation d'une fonction à une seule variable $\rho \in [0, 1]$, définie comme suit :

$$H_k(\rho) = \sum_{i=1}^{p} \bar{x}_i^2 \varphi^\star\left(\frac{L^i_k(\rho)}{-\bar{x}_i}\right) + G_k(\rho), \tag{4.23}$$

où $L_k(\rho) := \rho a^1_k + (1-\rho) a^2_k \in \mathbb{R}^p$, L^i_k est la $i$ème composante de la fonction L_k, et $G_k(\rho) = \rho b^1_k + (1-\rho) b^2_k$.

Proposition 4.3.1 *Pour la paire de problème (P_k)-(D_k) et pour $\bar{x} \in I\!R^p_{++}$, on a les relations suivantes :*

(a) $m_k(\bar{x}) + p_k(\bar{x}) = 0$;

(b) $p_k(\bar{x}) = \inf_{\rho \in [0,1]} H_k(\rho)$, et l'infinimum est atteint en un certain ρ_k ;

(c) La solution optimale x_k de (P_k) est donnée par

$$x_k^i = \bar{x}_i(\varphi^\star)' \left(\frac{L_k^i(\rho_k)}{-\bar{x}_i} \right) \qquad i = 1, \cdots, p.$$

Preuve :

Premièrement, un calcul direct montre que $\Phi^\star(u) = \sum_{i=1}^p \bar{x}_i^2 \varphi^\star(\bar{x}_i^{-1} u_i)$. Alors par le Théorème 3.1.1 nous avons que dom $\varphi^\star = \mathbb{R}$, il suit que dom $\Phi^\star = \mathbb{R}^p$, et par le Théorème de dualité de Fenchel ([8] p 327), (a) suit et le minimum est atteint en (D_k). Parce que la conjuguée $t^\star$ d'une fonction affine $t(x) = \langle a, x \rangle - b$ est donnée par la formule

$$t^\star(v) = \begin{cases} b & \text{si } v = a, \\ +\infty & \text{sinon} \end{cases}$$

alors par [8] (Théorème 16.5), il suit que pour chaque $u \in$ dom $(\psi^k)^\star$, nous avons

$$(\psi^k)^\star(u) = G_k(\rho), \quad \text{où } \rho \in [0,1] \text{est tel que } u = L_k(\rho).$$

Maintenant, $u \in$ dom $(\psi^k)^\star$ si et seulement si $u \in \mathrm{conv}(a_k^1, a_k^2)$. De plus, (b) est une conséquence de (4.22), et pour obtenir l'assertion (c), nous savons que ρ_k est minimum de $H_k(\rho)$ sur $[0,1]$ et $u_k = L_k(\rho_k)$ est solution de (D_k), donc en appliquant [8] (Théorème 31.3), nous avons que la solution x_k de (P_k) vérifie $x_k \in \partial\Phi^\star(-u_k)$. Et donc nous avons que $x_k^i = \bar{x}_i(\varphi^\star)' \left(\frac{L_k^i(\rho)}{-\bar{x}_i} \right)$ $i = 1, \cdots, p$. □

Chapitre 5

Les méthodes intérieures de descente du ϵ-sous gradient.

Dans ce chapitre, nous allons établir la convergence de plusieurs méthodes basées sur le sous-gradient pour le problème avec contraintes (P). Ensuite nous décrirons quelques applications et exemples pour illustrer notre résultat.

5.1 Description.

Nous considérons toujours le problème (P) défini précédemment, avec les mêmes hypothèses et les mêmes notations. Quand l'ensemble des contraintes C est l'ensemble $\mathbb{R}^p$, il est possible de montrer que la méthode proximale classique, et les méthodes de faisceaux classiques peuvent être considérées comme des méthodes de descente ϵ-sous gradient décrites comme suit.

On commence avec un $x_0 \in \mathbb{R}^p$ et on génère une succession de points à l'aide des formules

$$x^k = x^{k-1} - \lambda_k g^{k-1}, \qquad g^{k-1} \in \partial_{\epsilon_k} f(x^{k-1}), \tag{5.1}$$

où λ_k est un paramètre de pas positif, $\epsilon_k \geq 0$ et $\partial_{\epsilon_k} f(x^{k-1})$ est le ϵ_k-sous-différentiel de f en x^{k-1}.

Pour obtenir la convergence, nous devons imposer une sorte de propriété de descente : il existe une constante $m \in (0, 1]$ telle que pour chaque k nous avons

$$f(x^k) \leq f(x^{k-1}) + m(\langle g^{k-1}, x^k - x^{k-1}\rangle - \epsilon_k). \tag{5.2}$$

Nous savons que l'itération (5.1) n'est rien d'autre que

$$x^k = \arg\min \left\{ \lambda_k \langle g^{k-1}, x\rangle + \frac{1}{2}\|x - x^{k-1}\|^2 : x \in \mathbb{R}^p \right\}.$$

Donc pour résoudre le problème (P), nous pouvons modifier ce modèle en utilisant la distance logarithmique-quadratique pour manipuler directement les contraintes en remplaçant seulement la formule (5.1), c'est-à-dire, nous considérons le schéma

$$x^0 \in \text{int } C; \quad x^k = \arg\min\{\lambda_k\langle g^{k-1}, x\rangle + d(x, x^{k-1})\}, \quad g^{k-1} \in \partial_{\epsilon_k} f(x^{k-1}) \tag{5.3}$$

Remarquer que l'existence et l'unicité de $x^k \in \text{int } C$ est garantie par la Proposition 3.2.2.

5.2 Théorème de convergence.

Théorème 5.2.1 *Soit $\{x^k\}$ une suite générée par (5.3) et satisfaisant (5.2), avec $\epsilon_k \geq 0$, $\lambda_k > 0$. Posons $\sigma_n = \sum_{k=1}^n \lambda_k$ et $\alpha_k = \langle g^{k-1}, x^{k-1} - x^k\rangle$. Supposons que $f_\star$ est finie ; alors*

(i)

$$\sum_{k=1}^{\infty} \alpha_k < \infty, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty, \quad \alpha_k \geq \frac{\theta}{\lambda_k}\|x^k - x^{k-1}\|^2 \geq 0. \tag{5.4}$$

En particulier les suites positives $\{\alpha_k\}$ et $\{\epsilon_k\}$ convergent vers zéro.

(ii) Pour n'importe quel $z \in C$,

$$\sigma_n(f(x^n) - f(z)) \leq \theta\|x^\star - x^0\|_A^2 + \sum_{k=1}^{n} \lambda_k(\alpha_k + \epsilon_k).$$

(iii) La suite $\{f(x^k)\}$ est décroissante et quand $\sigma_n \to \infty$, elle converge vers $f_\star$.

(iv) Si $\sum_{k=1}^{\infty} \lambda_k(\alpha_k + \epsilon_k) < \infty$ (ce qui est vrai si λ_k est borné supérieurement), et si l'ensemble des solutions $X_\star$ est non vide alors $\{x^k\}$ converge vers une solution optimale de (P) quand $\sigma_n \to \infty$.
De plus, si $C = I\!R_+^p$ et $\lambda_k \geq \lambda_\star > 0$, alors

$$\lim_{k\to\infty} \inf g_i^k \geq 0, \qquad \lim_{k\to\infty} g_i^k x_i^k = 0, \quad \forall i. \tag{5.5}$$

(v) Supposons que λ_k est bornée supérieurement par un $\lambda^\star > 0$ et soit $x^\star \in X_\star$. Alors nous avons l'estimation globale

$$f(x^n) - f_\star \leq K\sigma_n^{-1}, \forall n, \tag{5.6}$$

avec $K = \theta\|x^\star - x^0\|_A^2 + \lambda^\star m^{-1}(f(x^0) - f_\star)$, qui se simplifie en $f(x^n) - f_\star \leq K(\lambda_\star n)^{-1}$ quand $\lambda_k \geq \lambda_\star > 0$.

Preuve :

(i) Par les conditions d'optimalité, l'équation (5.3) est équivalente à

$$\lambda_k g^{k-1} + \nabla_x d(x^k, x^{k-1}) = 0, \quad g^{k-1} \in \partial_{\epsilon_k} f(x^{k-1}), \tag{5.7}$$

et avec $x^0 \in \operatorname{int} C$, il suit $x^k \in \operatorname{int} C$ pour tout k. De la formule (3.20) avec $z = y = x^{k-1}$, nous obtenons alors

$$\lambda_k \alpha_k \underbrace{=}_{(5.7)} \langle \nabla_x d(x^k, x^{k-1}),\, x^k - x^{k-1} \rangle \underbrace{\geq}_{3.20} \theta \|x^k - x^{k-1}\|_A^2 \geq 0.$$

De plus, de (5.2) nous obtenons

$$0 \leq m(\alpha_k + \epsilon_k) \leq f(x^{k-1}) - f(x^k), \tag{5.8}$$

ce qui montre que $\{f(x^k)\}$ est décroissante. Sommant (5.8) pour $k = 1, \ldots, n$ il suit que

$$m \sum_{k=1}^{n} (\alpha_k + \epsilon_k) \leq f(x^0) - f(x^n) \leq f(x^0) - f_\star,$$

et donc (i) est prouvé.

(ii) Comme $g^{k-1} \in \partial_{\epsilon_k} f(x^{k-1})$, alors pour n'importe quel z nous avons

$$\begin{aligned} f(z) - f(x^{k-1}) + \epsilon_k &\geq \langle g^{k-1},\, z - x^{k-1} \rangle \\ &= \langle g^{k-1},\, z - x^k \rangle + \langle g^{k-1},\, x^k - x^{k-1} \rangle \\ &= -\frac{1}{\lambda_k} \langle z - x^k,\, \nabla_x d(x^k, x^{k-1}) \rangle - \alpha_k \\ &\underbrace{=}_{(3.20)} \frac{\theta}{\lambda_k} (\|z - x^k\|_A^2 - \|z - x^{k-1}\|_A^2) - \alpha_k, \end{aligned}$$

où la dernière inégalité est obtenue par (3.20) avec $x = x^k$, $y = x^{k-1}$. Il suit alors que

$$\lambda_k (f(x^{k-1}) - f(z)) \leq \theta(\|z - x^{k-1}\|_A^2 - \|z - x^k\|_A^2) + \lambda_k(\alpha_k + \epsilon_k).$$

Posons $z = x^\star$; parce que $f(x^k) \leq f(x^{k-1})$, nous obtenons

$$\lambda_k (f(x^k) - f_\star) \leq \theta(\|x^\star - x^{k-1}\|_A^2 - \|x^\star - x^k\|_A^2) + \lambda_k(\alpha_k + \epsilon_k), \tag{5.9}$$

et en sommant sur $k = 1, \ldots, n$, parce que $f(x^n) \leq f(x^k)\, \forall k \leq n$, nous obtenons

$$\sigma_n (f(x^n) - f_\star) \leq \theta \|x^\star - x^0\|_A^2 + \sum_{k=1}^{n} \lambda_k(\alpha_k + \epsilon_k). \tag{5.10}$$

(iii) Maintenant, divisant les deux membres de la dernière égalité par σ_n, en passant à la limite $n \to \infty$, puisque $\sigma_n \to \infty$ et α_k, $\epsilon_k \to 0$ et invoquant le Lemme 3.2.2, (iii) suit. En effet, $b_n = \sigma_n^{-1} \sum_{k=1}^{n} \lambda_k a_k$ où $a_k = \alpha_k + \epsilon_k \to 0$. Donc si $\sigma_n \to \infty$ alors $b_n \to 0$. Et donc,

$$(f(x^n) - f_\star) \leq \underbrace{\theta\sigma_n^{-1}\|x^\star - x^0\|_A^2}_{\to 0} + \underbrace{\sigma_n^{-1}\sum_{k=1}^{n}\lambda_k(\alpha_k + \epsilon_k)}_{\to 0}.$$

(iv) De (5.9), il suit que pour chaque solution optimale $x^\star$ nous avons

$$\|x^\star - x^k\|_A^2 \leq \|x^\star - x^{k-1}\|_A^2 + \mu_k$$

avec $\mu_k = \theta^{-1}\lambda_k(\alpha_k + \epsilon_k)$. Et par les hypothèses dans (iv), nous avons de (i) que $\sum_{k=1}^{\infty} \mu_k < +\infty$, et donc en utilisant le Lemme 3.2.1, il suit que la suite $\{\|x^\star - x^k\|_A^2\}$ converge, et d'où la suite $\{x^k\}$ converge vers un x_∞ solution optimale de (P).
Maintenant, supposons que $C = \mathbb{R}_+^p$. Alors par le Théorème 3.2.1 nous obtenons

$$\liminf_{k\to\infty} g_i^k \geq 0, \quad \lim_{k\to\infty} g_i^{k-1} x_i^k = 0, \quad \forall i. \tag{5.11}$$

Cependant, parce que $\{\alpha_k\}$ converge vers 0, il suit que

$$\lim_{k\to\infty} g_i^k x_i^k = 0, \quad \forall i$$

(v) Parce que $\lambda_k \leq \lambda^\star \forall k$, alors effectuant la somme sur (5.8) de $k = 1, \ldots, n$, il suit que

$$\sum_{k=1}^{n} \lambda_k(\alpha_k + \epsilon_k) \leq m^{-1}\lambda^\star(f(x^0) - f_\star).$$

Combinant cette dernière inégalité avec (5.10) nous obtenons

$$f(x^n) - f_\star \leq \sigma_n^{-1}\left\{\theta\|x^\star - x^0\|_A^2 + \lambda^\star m^{-1}(f(x^0) - f_\star)\right\} := \sigma_n^{-1}K,$$

ce qui prouve la première partie de (v). De plus, avec $\lambda_k \geq \lambda_\star > 0$, il suit que $f(x^n) - f_\star \leq K(n\lambda_\star)^{-1}$. □

5.3 Applications et exemples.

5.3.1 La méthode logarithmique-quadratique proximale exacte.

Nous considérons la suite $\{x^k\}$ construite avec la méthode LQP par les équations (3.21) et (3.22) avec $\epsilon_k = 0$. Commençant au point $x^0 \in \text{int } C$, la suite est uniquement définie et reste dans int C.

De plus, par la formule de transport (voir la Proposition 1.3.2) l'équation $g^k \in \partial f(x^k)$ est équivalente à

$$g^k \in \partial_{\epsilon_k^\star} f(x^{k-1}) \quad \text{avec } \epsilon_k^\star = f(x^{k-1}) - f(x^k) + \langle g^k,\, x^k - x^{k-1}\rangle.$$

En outre, parce que $g^k \in \partial f(x^k)$, il suit que $\epsilon_k^\star \geq 0$.
Maintenant, parce que (3.21) et (3.22) ne sont rien d'autre que l'équation (5.7), il suit qu'elles sont équivalentes à (5.3).
Finalement, avec $m = 1$ et avec $\epsilon_k^\star$ définit ci-dessus, l'inégalité (5.2) devient une égalité, et la suite $\{x^k\}$ construite par la méthode proximale logarithmique-quadratique satisfait (5.2) et (5.3).

5.3.2 Une méthode de faisceaux intérieure.

Considérons l'algorithme de faisceaux IBA décrit dans le chapitre précédent avec $0 < \lambda_\star \leq \lambda_k \leq \lambda^\star$ et on suppose aussi avoir les hypothèses A. Nous utilisons ici seulement la partie (a) du Théorème 4.2.2 qui dit que l'algorithme est bien défini. Supposons aussi que nous choisissons la première règle de faisceaux ; c'est-à-dire, ψ^k est définie par la formule (4.7) telle que par (4.9) $\psi^{n(k)}(y^k) = f(y^k)$.
Nous allons maintenant prouver que $\{y^k\}$ satisfait (5.2) et (5.3).
Par le Théorème 4.2.2 partie (a), parce que $d^k \in \partial\psi^{n(k)}(y^{k+1})$ nous avons

$$f(z) \underbrace{\geq}_{H1} \psi^{n(k)}(z) \geq \psi^{n(k)}(y^{k+1}) + \langle d^k,\, z - y^{k+1}\rangle = f(y^k) + \langle d^k,\, z - y^k\rangle - \epsilon_k,\ \forall z \in \mathbb{R}^p$$

avec

$$\epsilon_k = (f(y^k) - \psi^{\psi n(k)}(y^k)) + \psi^{n(k)}(y^k) - \psi^{n(k)}(y^{k+1}) - \langle d^k,\, y^k - y^{k+1}\rangle \quad (5.12)$$

Le premier terme du coté droit de (5.12) vaut zéro et comme $d^k \in \partial\psi^{n(k)}(y^{k+1})$ par la partie (a) du Théorème 4.2.2, il suit que $\epsilon_k \geq 0$ et que $d^k \in \partial_{\epsilon_k} f(y^k)$.
De plus, cette partie du Théorème 4.2.2 nous dit que $d^k = -\lambda_k^{-1}\nabla_x d(y^{k+1}, y^k)$ ce qui signifie que (5.3) est satisfait. Finalement de (4.18), (5.12), et parce que $\psi^{n(k)}(y^k) = f(y^k)$, nous obtenons

$$f(y^k) - f(y^{k+1}) \geq \eta(f(y^k) - \psi^{n(k)}(y^{k+1})) \geq \eta[\langle d^k,\, y^k - y^{k+1}\rangle + \epsilon_k],$$

et donc (5.2) est satisfait.

5.3.3 Une méthode de projection du gradient intérieur.

Ici, nous considérons le problème (P) avec f continûment différentiable sur C, et nous proposons la méthode suivante.

Algorithme 5.3.1 *Prendre* $m \in (0, 1)$. *Commencer avec* $x^0 \in int\ C$, $\lambda_0 > 0$ *et générer les suites* $\{x^k,\, \lambda_k\}$, *satisfaisant*

$$x^k := x^{k-1}(\lambda_k) = arg\ min_x\{\lambda_k\langle\nabla f(x^{k-1}),\, x\rangle + d(x, x^{k-1})\}, \quad (5.13)$$

avec λ_k jouant le rôle de longueur de pas tel que

$$f(x^{k-1}(\lambda_k)) \leq f(x^{k-1}) + m\langle \nabla f(x^{k-1}),\, x^{k-1}(\lambda_k) - x^{k-1}\rangle. \tag{5.14}$$

Par la Proposition 3.2.2, nous remarquons premièrement que l'existence d'un unique $x^{k-1}(\lambda_k) \in \text{int } C$ est garantie.

L'algorithme proposé ajuste la méthode générale décrite dans la section 5.1 avec $g^k = \nabla f(x^{k-1})$ et $\epsilon_k = 0$. Pour appliquer le Théorème 5.2.1 et obtenir le résultat de convergence, nous avons besoin d'établir que la suite $\{\lambda_k\}$ satisfait

$$\lambda_\star \leq \lambda_k \leq \lambda^\star, \tag{5.15}$$

pour les réels $\lambda^\star$, $\lambda_\star > 0$.

Nous allons décrire deux types de règles de pas satisfaisant (5.15).

Règle du pas fixe.
Nous supposons que ∇f est Lipschitz sur C ; c'est-à-dire, il existe $L > 0$ tel que

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x,\, y \in C. \tag{5.16}$$

Soit $\lambda_{\min}(A^T A)$ la plus petite valeur propre de la matrice symétrique définie positive $A^T A$, posons $\theta^\star = \lambda_{\min}(A^T A)\theta$, où θ est défini par (3.4). Soit aussi $\epsilon \in]0,\, 1[$, posons $\lambda^\star := 2\epsilon\theta^\star L^{-1}$, et choisissons $\lambda_\star \in]0,\, \lambda^\star[$. Maintenant, prenons $\lambda_\star \leq \lambda_k \leq \lambda^\star$; alors la méthode de projection du gradient intérieur se réduit à l'algorithme ci-dessous.

Algorithme 5.3.2 *Commencer au point $x^0 \in$ int C et construire la suite $\{x^k\}$ avec la formule*

$$x^k = arg\ min\{\lambda_k\langle \nabla f(x^{k-1}),\, x\rangle + d(x,\, x^{k-1})\}. \tag{5.17}$$

Nous devons seulement vérifier que l'inégalité (5.14) suit.

Pour se faire nous avons besoin du célèbre Lemme de descente que voici :

Lemme 5.3.1 *Si $f : I\!R^n \to I\!R$ est continûment différentiable et satisfait la propriété suivante*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in I\!R^n,$$

où L est un scalaire, alors

$$f(x + y) \leq f(x) + y'\nabla f(x) + \frac{L}{2}\|y\|^2.$$

Preuve :
Soit t un paramètre scalaire et soit $g(t) = f(x+ty)$. Alors grâce à la règle en chaîne nous avons $(dg/dt)(t) = y'\nabla f(x+ty)$. Maintenant

$$\begin{aligned}
f(x+y) - f(x) &= g(1) - g(0) = \int_0^1 \frac{dg}{dt}(t)dt = \int_0^1 y'\nabla f(x+ty)dt \\
&\leq \int_0^1 y'\nabla f(x)dt + \left| \int_0^1 y'(\nabla f(x+ty) - \nabla f(x))dt \right| \\
&\leq \int_0^1 y'\nabla f(x)dt + \int_0^1 \|y\| \cdot \|\nabla f(x+ty) - \nabla f(x)\|dt \\
&\leq y'\nabla f(x) + \|y\| \int_0^1 Lt\|y\|dt \\
&= y'\nabla f(x) + \frac{L}{2}\|y\|^2.
\end{aligned}$$

□

Parce que ∇f est Lipschitz, nous pouvons utiliser le Lemme et nous obtenons que

$$f(x^k) \leq f(x^{k-1}) + \langle \nabla f(x^{k-1}),\, x^k - x^{k-1}\rangle + \frac{L}{2}\|x^k - x^{k-1}\|^2. \tag{5.18}$$

De plus, parce que (5.17) est équivalent à $\lambda_k \nabla f(x^{k-1}) + \nabla_x d(x^k,\, x^{k-1}) = 0$, et parce que par la Proposition 3.2.1 nous avons

$$\langle \nabla_x d(x^k,\, x^{k-1}), x^{k-1} - x^k\rangle \leq -\theta\|x^k - x^{k-1}\|_A^2 \leq -\theta^\star\|x^k - x^{k-1}\|^2,$$

il suit de (5.18) que

$$f(x^k) \leq f(x^{k-1}) + \langle x^k - x^{k-1},\, \nabla f(x^{k-1})\rangle(1 - \frac{L\lambda_k}{2\theta^\star}),$$

et avec $\lambda^\star = (2\epsilon\theta^\star)/L$ nous obtenons l'inégalité désirée

$$f(x^k) \leq f(x^{k-1}) + \langle x^k - x^{k-1},\, \nabla f(x^{k-1})\rangle(1-\epsilon).$$

Dans le cas général, le calcul de x^k dans (5.17) n'est pas donné par une formule explicite et doit être produit par un algorithme. Cependant, parce que ce calcul consiste à la minimisation d'une somme d'une fonction linéaire et d'une fonction logarithmique quadratique, et grâce aux propriétés des fonctions logarithmiques quadratiques (en particulier la self concordance et la forte convexité), ceci peut être fait trés efficacement en utilisant une méthode de type Newton.

Remarque 5.3.1
Si f est quadratique convexe, c'est-à-dire, $f(x) = 2^{-1}\langle Qx, x\rangle + \langle c, x\rangle$, alors f est Lipschitz sur tout l'ensemble avec $L = \|Q\|$.

Pour illustrer l'importance de la formule (5.5) dans ce contexte, nous considérons le programme quadratique convexe :

$$(PQ) \qquad \alpha = \inf\{g(y) | By \geq d\},$$

avec $g(y) = \frac{1}{2}\langle Qy, y\rangle + \langle q, y\rangle$, Q une matrice $m \times m$ définie positive, et $q \in \mathbb{R}^m$, B une matrice $p \times m$.

Il est bien connu que le dual de Lagrange est donné par

$$(DP) \qquad \beta = \inf\{f(x) | x \in \mathbb{R}^p_+\},$$

avec $f(x) = g^\star(B^T x) - \langle d, x\rangle$, $g^\star(z) = \frac{1}{2}\langle z - q, Q^{-1}(z - q)\rangle$.

Nous supposons seulement que l'ensemble admissible est non vide. Alors le problème (PQ) admet une solution unique $\bar{y}$, $\alpha = -\beta$, et l'ensemble optimal $X_\star$ de (DP) est non vide.

Soit $\{x^k\}$ est une suite donnée explicitement par la formule (4.18), avec $\lambda_k \in]\lambda_\star, \lambda^\star[$, $\lambda^\star = 2\epsilon\theta \|BQ^{-1}B^T\|^{-1}$ et posons $y^k = \nabla g^\star(B^T x^k)$ avec $\nabla g^\star(z) = Q^{-1}(z - q)$. Alors la suite $\{(x^k, y^k)\}$ converge vers $(\bar{x}, \bar{y})$ avec $\bar{x} \in X_\star$. En effet, utilisant le Théorème 5.2.1 nous obtenons

$$\lim_{k\to\infty} x^k = \bar{x} \in X_\star, \quad \lim_{k\to\infty} \langle \nabla f(x^k), x^k\rangle = 0, \quad \liminf_{k\to\infty} \frac{\partial f}{\partial x_i}(x^k) \geq 0 \, \forall i.$$

Parce que $\nabla f(x^k) = B\nabla g^\star(B^T x^k) - d$, il suit que

$$\lim_{k\to\infty} y^k = y_\infty := \nabla g^\star(B^T \bar{x}), \quad By_\infty \geq d, \quad \langle By_\infty - d, \bar{x}\rangle = 0.$$

Finalement parce que $y_\infty = \nabla g^\star(B^T \bar{x}) \Leftrightarrow B^T \bar{x} = (\nabla g^\star)^{-1}(y_\infty) = \nabla g(y_\infty)$, il suit par les conditions d'optimalité que $y_\infty = \bar{y}$.

Une généralisation de la règle d'Armijo.

Pour simplifier la présentation, nous considérons ici le cas $C = \mathbb{R}^p_+$. Nous commençons avec la Proposition suivante qui sera un résultat clé.

Proposition 5.3.1 *Soit $c \in I\!R^p$ avec $c \neq 0$, et soit $x \in I\!R^p_{++}$. Pour chaque $\lambda > 0$, $x(\lambda)$ est l'unique point dans $I\!R^p_{++}$ satisfaisant*

$$x(\lambda) := arg \min_u \{\lambda\langle c, u\rangle + d(u, x)\}. \tag{5.19}$$

Alors,

(a) $\theta\|x - x(\lambda)\|^2 \leq \lambda\langle x - x(\lambda), c\rangle$.

(b) $\|(x - x(\lambda))/\lambda\| \leq \theta^{-1}\|c\|$.

(c) $\lim_{\lambda\to 0^+}(x - x(\lambda))/\lambda = (2\theta)^{-1}c$.

Preuve :

L'existence d'un unique $x(\lambda) \in \mathbb{R}^p_{++}$ est garantie par la Proposition 3.2.2. Premièrement, nous remarquons que (b) est une conséquence immédiate de (a). Prouvons (a). Avec les conditions d'optimalité pour (5.19), nous avons $\lambda c = -\nabla_u d(x(\lambda), x)$ tel que

$$\lambda\langle x - x(\lambda), c\rangle = \langle x(\lambda) - x, \nabla_u d(x(\lambda), x)\rangle,$$

et par (3.20), nous obtenons l'inégalité désirée (a). Pour prouver (c), nous observons d'abord que

$$c_i = -\lambda^{-1}\frac{\partial}{\partial u_i}d(u, x)|_{u=x(\lambda)} = -\lambda^{-1}x_i\varphi'(x_i^{-1}x_i(\lambda)), \quad \forall i = 1, \ldots, p,$$

avec φ définie en (3.7). Rappelant que $2\theta = \nu + \mu$, il est facile de vérifier que

$$2\theta(1 - t^{-1}) \leq \varphi'(t) \leq 2\theta(t - 1) \quad \forall t > 0,$$

et d'où il suit que pour chaque $i = 1, \ldots, p$, nous avons que

$$2\theta(\frac{x_i - x_i(\lambda)}{\lambda}) \leq c_i \leq 2\theta(\frac{x_i - x_i(\lambda)}{\lambda})\frac{x_i}{x_i(\lambda)}. \tag{5.20}$$

Maintenant de (a) et (b), il suit que $x_i(\lambda) \to x_i > 0$ quand $\lambda \to 0^+$, et la suite $\{(x - x(\lambda))/\lambda\}_{\lambda>0}$ est bornée. Et par conséquent, en passant à la limite quand $\lambda \to 0^+$ dans (5.20), nous obtenons (c). □

Nous pouvons maintenant définir une règle de type Armijo pour λ_k. Soit $\beta \in (0, 1), m \in (0, 1)$, et $s > 0$ un scalaire choisit. Alors, posons $\lambda_k = \beta^{j_k}s$ où j_k est le premier entier non négatif j tel que

$$f(x^{k-1}(\beta^j s)) - f(x^{k-1}) \leq m\langle\nabla f(x^{k-1}), x^{k-1}(\beta^j s) - x^{k-1}\rangle. \tag{5.21}$$

Proposition 5.3.2 *Supposons que $\nabla f(x^{k-1}) \neq 0$. Alors, λ_k est bien défini et satisfait $\lambda_k \leq s$. De plus, si ∇f est Lipschitz sur C, c'est-à-dire, satisfait (5.16), alors il existe $\lambda_\star > 0$ tel que $\lambda_k \geq \lambda_\star$, $\forall k$.*

Preuve :

Par définition, si λ_k existe, nous avons $\lambda_k \leq s$. Nous devons donc vérifier que λ_k existe. Pour simplifier les notations, nous posons $x := x^{k-1}, c := \nabla f(x^{k-1})$. Supposons que (5.21) n'est pas vérifiée. Donc nous avons

$$f(x(\beta^j s)) - f(x) > m\langle x(\beta^j s) - x, c\rangle, \quad \forall j \in \mathbb{N}, \tag{5.22}$$

où $x(\lambda)$ est défini par (5.19). Invoquant le Théorème des accroissement finis, il existe $z_j \in (x, x(\beta^j s))$ tel que

$$\langle \nabla f(z_j), \frac{x(\beta^j s) - x}{\beta^j s}\rangle > m\langle \frac{x(\beta^j s) - x}{\beta^j s}, c\rangle \quad \forall j \in \mathbb{N}.$$

Cependant, par la Proposition 5.3.1 (b), il suit que $z_j \to x$, quand $j \to \infty$. Donc, utilisant (c) de la même proposition et passant à la limite dans l'inégalité ci-dessus, nous obtenons $(2\theta)^{-1}(1-m)\|c\|^2 \leq 0$, tel que $c = 0$, est une contradiction.
Maintenant, prouvons, que quand ∇f est Lipschitz, nous avons

$$\lambda_k \geq \lambda_\star := \min\{\frac{2\beta(1-m)\theta}{L}, s\} > 0.$$

Comme pour le cas de la règle du pas fixe, nous pouvons utiliser le Lemme de descente, et nous obtenons

$$f(x(\lambda)) - f(x) \leq \langle c, x(\lambda) - x\rangle(1 - \frac{L\lambda}{2\theta}), \quad \forall \lambda > 0,$$

et donc (5.21) suit pour tout $j \in \mathbb{N}$ tel que $\beta^j s \leq (2\theta(1-m))/L$. Cependant, parce que, par définition, si $j_k \neq 0, \lambda_k\beta^{-1}$ ne satisfait pas (5.21), alors $\lambda_k\beta^{-1} > (2\theta(1-m))/L$ et il suit que $\lambda_k \geq \lambda_\star, \forall k$. □

5.3.4 Une méthode de gradient projeté intérieure sur un ensemble polyédral défini par égalités.

Considérons le problème suivant :

$$\inf\{f(x) : x \in V \cap \mathbb{R}^p_+\},$$

où $V = \{x : Ex = d\}$, avec $d \in \mathbb{R}^m$ et E est une matrice $(m \times p)$.
Nous supposons aussi que f est continûment différentiable et ∇f est Lipschitz sur $V \cap \mathbb{R}^p_+$, avec une constante de Lipschitz L.

Nous observons d'abord que le problème ci-dessus peut être reformulé de façon équivalente par le problème (P) qui suit :

$$f_\star = \min\{f_0(x) \mid x \in \mathbb{R}^p_+\} \quad \text{avec } f_0 = f + \delta_V.$$

Avec cette formulation, nous pouvons appliquer ce que nous avons vu précédemment avec la fonction objectif non-différentiable f_0 et l'ensemble de contrainte devient $\mathbb{R}^p_+$.

Soit $\epsilon \in]0,1[$ et $\lambda^\star = 2\epsilon\theta L^{-1}$. Choisir $\lambda_\star \in]0,\lambda^\star[$, et prendre $\lambda_k \in]\lambda_\star,\lambda^\star[$. Pour résoudre (P), nous proposons la méthode intérieure suivante :

Algorithme 5.3.3 *Commencer à partir du point $x_0 \in V \cap \mathbb{R}^p_{++}$ et construire la suite $\{x^k\} \subset V \cap \mathbb{R}^p_{++}$ comme suit :*
soit

$$x_i^k(\mu) = x_i^{k-1}(\varphi^\star)'(-\lambda_k \frac{d_i^{k-1}}{x_i^{k-1}}), \quad \forall i, \quad \textit{avec } d^{k-1} = \nabla f(x^{k-1}) + E^T\mu, \tag{5.23}$$

et μ^k satisfait

$$\sum_{i=1}^{p} e_{ji} x_i^k(\mu^k) = d_j, \quad \forall j. \tag{5.24}$$

Alors,

$$x^k = x^k(\mu^k). \tag{5.25}$$

Comme nous pouvons le voir, μ^k existe, x^k est défini uniquement et reste dans $V \cap \mathbb{R}^p_{++}$. En effet, (5.23) est équivalent à l'équation

$$x^k(\mu) = \arg\min\{\lambda_k(\langle \nabla f(x^{k-1}) + E^T\mu, x\rangle) + d(x, x^{k-1}) : x \in \mathbb{R}^p\}, \tag{5.26}$$

qui, comme prouvé dans la Proposition 3.2.2, est uniquement défini et reste dans $\mathbb{R}^p_{++}$. Alors, si nous considérons le problème

$$(P_k) \ \arg\min\ \{\lambda_k\langle \nabla f(x^{k-1}), x\rangle + d(x, x^{k-1}) : x \in V\},$$

les conditions d'optimalité assurent que la solution unique de (P_k) est le point x^k qui satisfait (5.23), (5.24), et que

$$\begin{aligned} \langle \nabla_x d(x^k, x^{k-1}), x^{k-1} - x^k\rangle &= \lambda_k\langle \nabla f(x^{k-1}) + E^T\mu^k, x^k - x^{k-1}\rangle \\ &= \lambda_k\langle \nabla f(x^{k-1}), x^k - x^{k-1}\rangle. \end{aligned} \tag{5.27}$$

Maintenant, prouvons que cet algorithme reste dans le cadre du mémoire ; pour cela, nous devons prouver que la suite $\{x^k\}$ convergeant vers la solution optimale de (P) satisfait l'estimation (5.6). Nous devons prouver que la suite $\{x^k\}$ satisfait (5.3) et (5.2) avec $\epsilon_k = 0, m = 1-\epsilon$. Posons $g^{k-1} = \nabla f(x^{k-1}) + E^T\mu^k$; évidemment $g^{k-1} \in \partial f_0(x^{k-1})$, ce qui implique avec (5.24) que (5.3) est satisfait. De plus, parce que ∇f est Lipschitz nous avons (5.18), et parce que de (3.20) nous avons

$$\langle \nabla_x d(x^k, x^{k-1}), x^{k-1} - x^k \rangle \leq -\theta \|x^k - x^{k-1}\|^2,$$

utilisant (5.27) et le fait que $f_0(x^k) = f(x^k)$, $f_0(x^{k-1}) = f(x^{k-1})$, nous obtenons

$$f_0(x^k) \leq f_0(x^{k-1}) + \langle g^{k-1}, x^k - x^{k-1} \rangle (1 - \frac{L\lambda_k}{2\theta}).$$

Alors, avec $\lambda^\star = 2\epsilon(\theta/L)$, nous avons

$$f_0(x^k) \leq f_0(x^{k-1}) + \langle g^{k-1}, x^k - x^{k-1} \rangle (1 - \epsilon),$$

ce qui n'est rien d'autre que (5.2).

5.4 Une estimation efficace améliorée.

Pour estimer l'efficacité potentielle de cet algorithme, il sera désirable d'améliorer l'estimation globale donnée par la formule (5.6). Celle-ci sera maintenant donnée par la méthode logarithmique quadratique sans erreur.

Pour la simplicité de l'exposé, nous considérons le cas du problème (P) avec $C = \mathbb{R}^p_+$. Alors, la méthode LQP sans erreur définie dans le chapitre 3 devient :

Commençant au point $x^0 \in \mathbb{R}^p_{++}$ et la suite $\{\lambda_k\} \subset \mathbb{R}_{++}$, on construit la suite $x^k \subset \mathbb{R}^p_{++}$ satisfaisant

$$\lambda_k g^k + \nabla_x d(x^k, x^{k-1}) = 0 \text{ avec } g^k \in \partial f(x^k), \tag{5.28}$$

où $d(x, y) = d_\varphi(x, y)$ est définie par (3.13), avec φ donnée par (3.7).

Lemme 5.4.1 *Soit $\nu > \mu > 0$ et θ défini en (3.4). Pour n'importe quelle $a, b \in \mathbb{R}^p_{++}$ et $c \in \mathbb{R}^p_+$, nous avons*

(i) $\langle c - b, \nabla_x d(b, a) \rangle \leq \mu \langle b - a, c - a \rangle + \nu \langle b - a, c - b \rangle$,

(ii) $\langle c - b, \nabla_x d(b, a) \rangle \leq \theta(\|c - a\|^2 - \|c - b\|^2) - ((\nu - \theta)/2)\|b - a\|^2$.

Preuve :

Identique à la preuve du Lemme 3.1.2 avec $\Phi'(a, b)$ remplacé par $\nabla_x d(b, a)$.
□

Théorème 5.4.1 *Soit $\{x_k\}$ une suite générée par la méthode LQP, et $\theta_n := \sum_{k=1}^{n} \lambda_k$. Supposons que l'ensemble des solutions optimales $X_\star$ est non vide, et $\sigma_n \to \infty$. Alors, le taux de convergence estimé est*

$$f(x^n) - f(x^\star) = o(\frac{1}{\sigma_n}),$$

c'est-à-dire, $\sigma_n(f(x^n) - f(x^\star)) \to 0$.
En particulier, si $\lambda_k \geq \lambda > 0$, nous avons le taux de convergence estimé $f(x^n) - f(x^\star) = o((\lambda n)^{-1})$.

Preuve :

Soit $\{x_k\}$ est une suite générée par la méthode LQP. Nous observons d'abord que

$$\lambda_k(f(x^k) - f(x)) \leq \langle x - x^k, \nabla_x d(x^k, x^{k-1})\rangle \tag{5.29}$$

$$f(x) + \lambda_k^{-1} d(x, x^{k-1}) \geq f(x^k) + \lambda_k^{-1} d(x^k, x^{k-1}), \quad \forall x \in \mathbb{R}^p_{++}, \tag{5.30}$$

$$\lambda_k(f(x^{k-1}) - f(x^k)) \geq d(x^k, x^{k-1}) \geq \frac{\nu}{2}\|x^k - x^{k-1}\|^2. \tag{5.31}$$

La première inégalité suit de (5.28) et de l'inégalité du sous-gradient pour la fonction convexe f.
La seconde inégalité vient de la première, de la définition de d et de la convexité de $d(\cdot, x^{k-1})$.
Et la dernière est juste un cas spécial de (5.30) avec $x = x^{k-1}$, puisque $d(x^{k-1}, x^{k-1}) = 0$. Et du fait que $d(u, v) = \mu d_\varphi(u, v) + (\nu/2)\|u - v\|^2 \geq \|u - v\|^2$, parce que $\mu d_h(\cdot, \cdot) \geq 0$, nous obtenons la seconde inégalité de (5.31).

Maintenant, par le Théorème 3.2.1, il suit que la suite $\{x^n\}$ converge vers la solution optimale $x^\star \in X_\star \neq \emptyset$, comme $\sigma_n \to \infty$. Notons $v_k := f(x^k) - f(x^\star) > 0$ pour chaque k avec $f(x^\star) \neq f(x^k)$. Notre objectif est de dérivé une inégalité récursive appropriée pour v_k. Appliquant (5.29) à $x = x^\star$ et le Lemme 5.4.1(i) avec $c = x^\star, a = x^{k-1}$, et $b = x^k$, nous obtenons

$$\begin{aligned}
\lambda_k(f(x^k) - f(x)) &\leq \mu\langle x^k - x^{k-1}, x^\star - x^{k-1}\rangle + \nu\langle x^k - x^{k-1}, x^\star - x^k\rangle \\
&\leq \|x^k - x^{k-1}\|\{\mu\|x^\star - x^{k-1}\| + \nu\|x^\star - x^k\|\} \\
&\leq (\mu + \nu)\|x^\star - x^{k-1}\|\|x^k - x^{k-1}\|,
\end{aligned}$$

où dans la seconde inégalité, nous utilisons $\|x^\star - x^k\| \leq \|x^\star - x^{k-1}\|, \forall k$, qui vient du Théorème 3.2.1(i). De plus, de la dernière inégalité ci-dessus, nous pouvons prouver

$$\|x^k - x^{k-1}\| \geq \frac{\lambda_k(f(x^k) - f(x^\star))}{(\mu + \nu)\|x^\star - x^{k-1}\|} = \frac{\lambda_k v_k}{2\theta\|x^\star - x^{k-1}\|}. \tag{5.32}$$

invoquant (5.31), nous avons aussi

$$\lambda_k(f(x^{k-1}) - f(x^\star)) - (f(x^k) - f(x^\star)) \geq \frac{\nu}{2}\|x^{k-1} - x^k\|^2,$$

qui combiné avec (5.32) donne

$$v_{k-1} - v_k \geq \eta_k v_k^2, \tag{5.33}$$

où nous avons défini

$$\eta_k := \frac{\lambda_k \nu}{8\theta^2\|x^\star - x^{k-1}\|^2} > 0, \quad \forall k. \tag{5.34}$$

Maintenant, appliquant (5.29) en $x = x^\star$ et le Lemme 5.4.1(ii) aux points $c = x^\star, a = x^{k-1}$, et $b = x^k$, il suit que

$$\lambda_k v_k \leq \theta\|x^\star - x^{k-1}\|^2. \tag{5.35}$$

En inversant l'inégalité (5.33), nous obtenons

$$\frac{1}{v_k - 1} \leq \frac{1}{v_k}(\frac{1}{1 + \eta_k v_k}), \quad \forall k = 1, \ldots \tag{5.36}$$

Utilisant (5.34) et (5.35), nous avons

$$\eta_k v_k = \frac{\lambda_k v_k}{\theta\|x^\star - x^{k-1}\|^2}\frac{\nu}{8\theta} \leq \frac{\nu}{8\theta} < \frac{1}{4}, \tag{5.37}$$

où dans la dernière inégalité nous utilisons la définition de θ (cfr (3.4)) et donc $(\nu/8\theta) = \frac{1}{4} - \frac{1}{8}\mu < \frac{1}{4}$. Maintenant il est facile de vérifier que $(1+t)^{-1} \leq 1 - \frac{4}{5}t$, $\forall t \in [0, \frac{1}{4}]$. Parce que par (5.37) nous avons $\eta_k v_k \in (0, \frac{1}{4})$, utilisant la dernière inégalité avec $t := \eta_k v_k$, nous obtenons de (5.36) que

$$\frac{1}{v_{k-1}} \leq \frac{1}{v_k}(1 - \frac{4}{5}\eta_k v_k), \tag{5.38}$$

et lorsque nous sommons sur $k = 1, \ldots, n$ nous obtenons

$$v_n \leq \frac{v_0}{(1 + \frac{4}{5}v_0 \sum_{k=1}^n \eta_k)}, \tag{5.39}$$

et en substituant par (5.34) cette inégalité se réduit à

$$v_n \leq \frac{10\theta^2\nu^{-1}}{\sum_{k=1}^n \lambda_k\|x^\star - x^{k-1}\|^{-2}}. \tag{5.40}$$

Multipliant (5.40) par $\sigma_n = \sum_{k=1}^n \lambda_k$, il suit que

$$\sigma_n(f(x_n) - f(x^\star)) \leq \frac{10\theta^2\nu^{-1}}{\sigma_n^{-1}\sum_{k=1}^n \lambda_k\|x^\star - x^{k-1}\|^{-2}}. \tag{5.41}$$

Cependant, parce que $\|x^{k-1} - x^\star\| \to 0$, nous avons $\|x^{k-1} - x^\star\|^{-2} \to \infty$ et par le Lemme 3.2.2 $\sigma_n^{-1} \sum_{k=1}^n \lambda_k \|x^\star - x^{k-1}\|^{-2} \to \infty$. De plus, comme le coté droit de (5.41) tends vers 0, nous avons que $f(x^n) - f(x^\star) = o((\sigma_n)^{-1})$ et nous avons prouvé le résultat désiré. $\square$

Maintenant comme une application, nous considérons les méthodes des multiplicateurs qui viennent de l'utilisation de la méthode LQP sur le dual d'un problème convexe à contrainte ordinaire :

$$(\tilde{P}) \quad \alpha = \inf\{f_0(x) : f_i(x) \leq 0, i = 1, \ldots, p\},$$

où $f_i : \mathbb{R}^n \to \mathbb{R}$ sont des fonctions convexes pour $i = 0, \ldots, p$.

Le problème dual classique associé au problème $(\tilde{P})$ est

$$(\tilde{D}) \quad \beta = \sup\{p(u) : u \in \mathbb{R}^m\},$$

avec

$$p(u) = \begin{cases} \inf\{L(x,y) : x \in \mathbb{R}^n\} & \text{si } u \geq 0, \\ +\infty & \text{sinon,} \end{cases}$$

et $L(x,y) = f_0(x) + \sum_{i=1}^p u_i f_i(x)$ est le Lagrangien usuel associé à $(\tilde{P})$.

Nous supposons que l'ensemble optimal de $(\tilde{P})$ est non vide et compact ce qui est équivalent à dire que

$$(f_i)_\infty(d) \leq 0 \quad \forall i = 0, \ldots, p \Rightarrow d = 0. \tag{5.42}$$

De plus, la condition de Slater est satisfaite. Sous les hypothèses ci-dessus, la fonction duale $-p(u)$ est propre, convexe et sci, $\alpha = \beta$, et l'ensemble dual optimal T de $(\tilde{D})$ est non vide et compact.

Pour $u > 0$, nous définissons

$$H(x, u, \lambda) = f_0(x) + \lambda^{-1} \sum_{i=1}^p u_i^2 \varphi^\star \left(\frac{\lambda f_i(x)}{u_i} \right), \tag{5.43}$$

où $\varphi^\star$ est la conjuguée de φ. Pour résoudre le problème $(\tilde{P})$, nous utilisons la méthode des multiplicateurs suivante :

Méthode des multiplicateurs (MM) :

Algorithme 5.4.1 *Soit* $u^0 \in \mathbb{R}^m_{++}$ *et* $\lambda_k \geq \lambda > 0$, $\forall k \geq 1$, *généré la suite* $\{x^k, u^k\}$ *satisfaisant*

$$x^k = \textit{arg min } \{H(x, u^{k-1}, \lambda_k) : x \in \mathbb{R}^n\}, \tag{5.44}$$

$$u_i^k = u_i^{k-1} (\varphi^\star)' \left(\frac{\lambda_k f_i(x^k)}{u_i^{k-1}} \right), \quad i = 1, \ldots, p. \tag{5.45}$$

Les preuves suivantes sont faites pour la méthode des multiplicateurs inexactes (IMM). Dans ce cas, l'équation (5.44) est remplacée par

$$x^k \approx \arg\min \{H(x, u^{k-1}, \lambda_k) : x \in \mathbb{R}^n\}, \tag{5.46}$$

où cette approximation signifie que pour $\epsilon_k \geq 0$, x^k doit satisfaire

$$L(x^k, u^k) \leq \inf_x L(x, u^k) + \epsilon_k = -p(u^k) + \epsilon_k. \tag{5.47}$$

Remarque 5.4.1 *Dans la méthode des multiplicateurs, c'est-à-dire, quand* $\epsilon_k = 0, \forall k,$ *(5.46) se réduit à* $x^k \in$ *argmin* H, *et avec (5.45) nous avons que* $\nabla_x H(x^k, u^{k-1}, \lambda_k) = \nabla_x L(x^k, u^k)$, *et (5.47) est satisfait.*

Proposition 5.4.1 *Soit* $\varphi \in \Phi$ *et supposons que (5.42) est satisfait. Alors,* $\forall \lambda > 0$, $u \in \mathbb{R}^p_{++}$ *l'ensemble des solutions argmin* $\{H(x, u, \lambda) : x \in \mathbb{R}^n\}$ *est non vide et compact.*

Preuve :

Par le Lemme 3.4.1, nous avons $\varphi^\star_\infty(-1) = 0$, $\varphi^\star_\infty(1) = +\infty$. Alors appliquant [4] Proposition 2.1 et utilisant l'homogénéité de $\varphi^\star_\infty$, nous obtenons

$$H_\infty(d) = \begin{cases} (f_0)_\infty(d) & \text{si } (f_i)_\infty(d) \leq 0, \quad \forall i = 1, \ldots, p. \\ +\infty & \text{sinon} \end{cases}$$

D'où, de (5.42) il suit que $H_\infty(d) > 0 \quad \forall d \neq 0$, et l'ensemble optimal de H est non vide et compact. □

Proposition 5.4.2 *Soit* $\{x^k, u^k\}$ *la suite générée par l'agorithme IMM et on définit* $F(x) := (f_1(x), \ldots, f_p(x))^T$. *Alors, la suite* $\{u^k, -F(x^k)\}$ *satisfait les relations (3.21) et (3.22) appliqué à la fonction duale p. De plus, nous avons que* $\nabla_x H(x^k, u^{k-1}, \lambda_k) = \nabla_x L(x^k, u^k)$.

Preuve :

Utilisant la définition de la fonction duale p nous avons

$$\langle F(x^k), u - u^k \rangle = L(x^k, u) - L(x^k, u^k) \geq -p(u) - L(x^k, u^k),$$

et puisque par (5.47) nous avons $L(x^k, u^k) \leq -p(u^k) + \epsilon_k$, il suit que

$$p(u) \geq p(u^k) + \langle -F(x^k), u - u^k \rangle - \epsilon_k, \forall u,$$

ce qui est équivalent à dire que

$$-F(x^k) \in \partial_{\epsilon_k} p(u^k). \tag{5.48}$$

Finalement, rappelant que par le Lemme 3.4.1 $(\varphi^\star)' = (\varphi')^{-1}$, la relation (5.45) est équivalente à

$$F(x^k) = \Phi'(u^{k-1}, u^k), \tag{5.49}$$

et d'où (5.48) et (5.49) prouvent le premier résultat.
De plus, utilisant (5.45) nous avons aussi

$$\begin{aligned}\nabla_x H(x^k, u^{k-1}, \lambda_k) &= \nabla f_0(x^k) + \sum_{i=1}^{p} u_i^{k-1} (\varphi^\star)'(\lambda_k f_i(x^k)/u_i^{k-1}) \nabla f_i(x^k), \\ &= \nabla f_0(x^k) + \sum_{i=1}^{p} u_i^k \nabla f_i(x^k) = \nabla_x L(x^k, u^k).\end{aligned}$$

□

Théorème 5.4.2 *Soit $\{x^k, u^k\}$ une suite générée par l'algorithme IMM, et considéront les hypothèses suivantes :*

(i) $\varphi \in \Phi_1$, *les conditions de Slater sont satisfaites,* $0 < \lambda < \lambda_k \leq \bar{\lambda}, \forall k, \sum \epsilon_k < \infty$,

(ii) $\varphi(t) = \mu h(t) + (\nu/2)(t-1)^2$, $h \in \Phi_2, \sum \epsilon_k < \infty, \sum \lambda_k \to +\infty, \sum \lambda_k \epsilon_k < \infty$.

Alors, sous (i) ou (ii) la suite $\{x^k, u^k\}$ est bornée et tous les points limites sont des solutions optimales de $(\tilde{P}) \times (\tilde{D})$. De plus, sous (ii), la suite duale $\{u^k\}$ converge globalement vers la solution de $(\tilde{D})$.

Preuve :

De la Proposition précédente, les relations (5.48) et (5.49) sont les relations de l'algorithme 3.2.1 utilisées pour résoudre le problème dual. Comme une conséquence du Théorème 3.2.1, nous obtenons que la suite $\{u^k\}$ est bornée, tous les points limites sont des solutions optimales dans le premier cas, et dans le second cas $\{u^k\}$ converge globalement vers la solution optimale du dual. De plus, dans les deux cas, les relations du Théorème 3.2.1 (iv) sont valables, d'où nous obtenons

$$\limsup_{k \to +\infty} f_i(x^k) \leq 0 \qquad \forall i = 1, \ldots, p, \tag{5.50}$$

$$\lim_{k \to \infty} f_i(x^k) u_i^k = 0 \qquad \forall i = 1, \ldots, p. \tag{5.51}$$

Puisque par (5.47) nous avons $L(x^k, u^k) \leq -p(u^k) + \epsilon_k$, alors par (5.51) nous obtenons

$$\limsup_{k \to \infty} f_0(x^k) \leq -\beta = \alpha. \tag{5.52}$$

Puisque l'ensemble optimal est borné, il suit de (5.50) et (5.52) que la suite $\{x^k\}$ est bornée. Soit x^∞ un point limite de $\{x^k\}$. Donc, puisque f_i est fermée, nous avons de (5.50)- (5.52) que $f_0(x^\infty) \leq \alpha$, $f_i(x^\infty) \leq 0$, $i = 1, \ldots, p$, et il suit que le point limite x^∞ est la solution optimale du primal. $\square$

Nous pouvons maintenant appliquer le Théorème 5.4.1 à la suite duale avec $f := -p$, et nous obtenons pour la suite duale le taux de convergence global estimé suivant :

$$p(u^\star) - p(u^n) = o(\sigma_n^{-1}).$$

En particulier, pour n'importe quel $\lambda_k \geq \lambda > 0$ nous avons $p(u^\star) - p(u^n) = o(\lambda_n^{-1})$.

Chapitre 6

Une méthode intérieure du sous-gradient.

Dans cette section, nous présentons une extension de la méthode du sous-gradient bien connu de Polyak.
Nous considérons le Problème (P) avec des contraintes positives ; c'est-à-dire, $C = \mathbb{R}^p_+$ et nous faisons les hypothèses suivantes.

Hypothèses B :
(a) $f_\star = \inf\{f(x) : x \in \mathbb{R}^p_+\} \in \mathbb{R}$.
(b) f continue sur $\mathbb{R}^p_+$.
(c) ∂f est borné sur l'ensemble des bornés de $\mathbb{R}^p_{++}$.

6.1 Motivation.

A l'itération k, ayant $x^k \in \mathbb{R}^p_+$ et la fonction estimée f^k_{lev} de $f_\star$, nous définissons pour chaque $g^k \in \partial f(x^k)$:

$$H_k^- := \{x \in \mathbb{R}^p : f(x^k) + \langle g^k, x - x^k\rangle \leq f^k_{\text{lev}}\}.$$

Fixons $\epsilon \geq 0$. L'algorithme du sous-gradient classique de Polyak consiste à construire une suite $\{x^k\}$ dans $\mathbb{R}^p_+$ tel que

$$x^{k+1} = P_{\mathbb{R}^p_+}(P_{H_k^-}(x^k)); \quad f^k_{\text{lev}} = f_\star + \epsilon, \tag{6.1}$$

où P_X est la projection euclidienne sur l'ensemble convexe fermé X. Il a été prouvé (Polyak 1969) que quand l'ensemble $\epsilon-$optimal est non vide, alors la suite converge vers un point x_∞ dans l'ensemble des $\epsilon-$optimal. Une difficulté de cette méthode dans le cas avec contrainte est qu'elle peut se comporter assez mal parce que la projection sur $\mathbb{R}^p_+$ est la source d'un comportement de zig-zag, donc cet algorithme n'est pas très performant.

6.2 L'algorithme : Analyse de la convergence.

Nous allons d'abord rappeler le concept de projection basé sur la distance logarithmique-quadratique. Nous définissons la projection logarithmique-quadratique sur $X \cap \mathbb{R}_+^n$, basé sur la distance d_φ via :
Pour chaque $y \in \mathbb{R}_{++}^p$,

$$E_X(y) := \arg\min\ \{d_\varphi(x, y) : x \in X \cap \mathbb{R}_+^p\}.$$

Lemme 6.2.1 *Soit X un ensemble convexe, fermé tel que $X \cap \mathbb{R}_{++}^p \neq \emptyset$. Alors,*

(i) Pour chaque $y \in \mathbb{R}_{++}^p$ la projection $E_X(y)$ existe, est unique et $E_X(y) \in X \cap \mathbb{R}_{++}^p$.

(ii) Pour n'importe quel $x \in X \cap \mathbb{R}_+^p$ nous avons

$$\alpha\|E_X(y) - y\|^2 \leq \|x - y\|^2 - \|x - E_X(y)\|^2,$$

où nous posons $\alpha := (\nu - \mu)(\nu + \mu)^{-1}$.

Preuve : Voir [2] Proposition 2.1 □

Maintenant, décrivons l'algorithme de base. Fixons $\epsilon \geq 0$. Soit $\{\epsilon_k\}$ une suite positive telle que $\sum_{k=1}^{\infty} \epsilon_k < +\infty$ et $\{f_{\text{lev}}^k\}$ une suite de nombres réels. Pour chaque $x^k \in \mathbb{R}_{++}^p$ nous supposons que nous savons calculer $g^k \in \partial f(x^k)$, et désignons

$$\begin{aligned}
S_\epsilon &= \{x \in \mathbb{R}_+^p : f(x) \leq f_\star + \epsilon\},\\
S_\epsilon^\star &= \{x \in \mathbb{R}_{++}^p : f(x) \leq f_\star + \epsilon\},\\
H_k &= \{x \in \mathbb{R}^p : f(x^k) + \langle g^k, x - x^k\rangle = f_{\text{lev}}^k\},\\
H_k^- &= \{x \in \mathbb{R}^p : f(x^k) + \langle g^k, x - x^k\rangle \leq f_{\text{lev}}^k\}.
\end{aligned}$$

Remarquer que pour $\epsilon > 0$, $S_\epsilon^\star$ est toujours non vide.

Nous donnons d'abord une extension de l'algorithme du sous-gradient de Polyak qui dépend de la suite $\{f_{\text{lev}}^k\}$. Après, nous donnerons une règle pour choisir f_{lev}^k à chaque étape, afin d'obtenir un algorithme implémentable pour résoudre (P).

Algorithme de base : Alg(f_{lev}^k).

Algorithme 6.2.1 *Commencer au point $x_0 \in \mathbb{R}_{++}^p$ et générer une suite itérativement $\{x^k\} \subset \mathbb{R}_{++}^p$ comme suit. Supposons qu'à l'itération k nous pouvons calculer $x^k > 0$ et $g^k \in \partial f(x^k)$.*
Si un des cas suivants se produit, alors **on arrête** *:*

(a) $g^k = 0$;

(b) $g^k \geq 0$ *et* $\langle g^k, x^k \rangle \leq \epsilon$;

(c) $x^k \in H_k^-$.

Sinon, calculer x^{k+1}*, satisfaisant :*

$$x^{k+1} \in I\!R^p_{++}, \quad x^{k+1} = E_{H_k}(x^k) + e_k, \ \textit{avec} \ \|e_k\| \leq \epsilon_k. \tag{6.2}$$

Théorème 6.2.1 *Supposons que l'ensemble optimal* S *est non vide quand* $\epsilon = 0$.

(a) Si au pas k nous avons que $g^k = 0$ *alors* x^k *est optimal.*
Si $g^k \geq 0$ *et* $\langle g^k, x^k \rangle \leq \epsilon$*, alors* $f(x^k) < f_\star + 2\epsilon$.
Si aucun des cas où l'agorithme s'arrête est satisfait, et si

$$H_k^- \cap I\!R^p_{++} \neq \emptyset, \tag{6.3}$$

alors, il existe un x^{k+1} *satisfaisant (6.2), et l'algorithme est bien défini.*

(b) Supposons que l'agorithme ne s'arrête pour aucun des pas, (6.3) est satisfait et il existe un x *tel que*

$$x \in \left(\cap_{k=0}^{\infty} H_k^-\right) \cap I\!R^p_{+}. \tag{6.4}$$

Alors la suite entière $\{\|x^k - x\|\}$ *converge vers la limite* $a(x) \geq 0$*, et nous avons que* $lim_{k\to\infty}\|x^k - E_{H_k}(x^k)\| = 0$.

Preuve :

(a) Si $g^k = 0$, évidement x^k est une solution optimale. Supposons que $g^k \geq 0$, $g^k \neq 0$, et $\langle g^k, x^k \rangle \leq \epsilon$, et soit $D_k = \{x \in \mathbb{R}^p : f(x^k) + \langle g^k, x - x^k \rangle \leq f(x^k) - \epsilon\}$. Nous avons alors que $D_k \cap \mathbb{R}^p_{++} = \emptyset$. En effet, il est équivalant de dire que pour $x \in D_k$, nous avons $\langle g^k, x - x^k \rangle \leq -\epsilon$, qui n'est pas satisfait pour $x > 0$. Maintenant,

- si $\epsilon = 0$, x^k est une solution optimale par les conditions de complémentarité. En effet, alors nous avons que $\langle g^k, x^k \rangle = 0$ et donc $\langle g^k, x^k - x \rangle \leq 0$ ce qui est équivalent à dire que x^k est une solution optimale.
- si $\epsilon > 0$, $S^\star_\epsilon$ est non vide et $S^\star_\epsilon \cap D_k = \emptyset$ car $S^\star_\epsilon \subseteq \mathbb{R}^p_{++}$ et $D_k \cap \mathbb{R}^p_{++} = \emptyset$.

Par conséquent, puisque f est convexe, pour n'importe quel $x \in S^\star_\epsilon$ nous avons

$$f_\star + \epsilon \geq f(x) \geq f(x^k) + \langle g^k, x - x^k \rangle > f(x^k) - \epsilon \text{ et } f(x^k) < f_\star + 2\epsilon.$$

Finalement, supposons que $g^k \neq 0$ et $g^k \geq 0$, $\langle g^k, x^k \rangle \leq \epsilon$ n'est pas satisfait, et que $x^k \notin H_k^-$ et (6.3) sont satisfaits. Alors, il existe un $x_k^\star \in$

$H_k \cap \mathbb{R}^p_{++} \cap [z, x^k]$, où $z \in H_k^- \cap \mathbb{R}^p_{++}$. Et donc, $H_k \cap \mathbb{R}^p_{++} \neq \emptyset$ et par le Lemme 6.2.1, $z_k := E_{H_k}(x^k)$ existe impliquant l'existance d'un point x^{k+1} qui satisfait (6.2).

(b) Maintenant supposons qu'à chaque pas, l'algorithme ne s'arrête pas, et que (6.3) est satisfait. Alors, utilisant le Lemme 6.2.1, nous obtenons que pour chaque x satisfaisant (6.4)

$$\alpha \|z^k - x^k\|^2 \leq \|x^k - x\|^2 - \|z^k - x\|^2, \tag{6.5}$$

ce qui implique que

$$\|z^k - x\| \leq \|x^k - x\|. \tag{6.6}$$

De plus,

$$\|x^{k+1} - x\|^2 = \|z^k - x + e_k\|^2 \leq \|z^k - x\|^2 + 2\epsilon_k \|z^k - x\| + \epsilon_k^2 = (\|z^k - x\| + \epsilon_k)^2.$$

la dernière égalité implique

$$\|x^{k+1} - x\| \leq \|z^k - x\| + \epsilon_k, \tag{6.7}$$

et utilisant (6.6) nous obtenons $\|x^{k+1} - x\| \leq \|x^k - x\| + \epsilon_k$. Donc, utilisant le Lemme 3.2.1, il suit que $\{\|x^k - x\|\}$ converge vers $a(x)$. Maintenant de (6.6) et (6.7) nous obtenons $\lim_{k\to\infty} \|z^k - x\| = a(x)$, et donc de (6.5) il suit que $\lim_{k\to\infty} \|x^k - E_{H_k}(x^k)\| = 0$. □

Maintenant, nous allons fournir deux règles pour choisir f^k_{lev} afin d'obtenir un algorithme implémentable.

Alg1 qui est obtenu à partir de $\mathbf{Alg}(f^k_{\text{lev}})$ avec $f^k_{\text{lev}} := f_\star + \epsilon$, $\forall k$.

Théorème 6.2.2 *Soit $S^\star_\epsilon = \{x \in \mathbb{R}^p_{++} : f(x) \leq f_\star + \epsilon\}$ non vide.*

(a) Si $x^k \in H_k^-$ alors $f(x^k) \leq f_\star + \epsilon$.

(b) **Alg1** *est bien défini.*

(c) Si **Alg1** *n'est pas arrêté, alors la suite $\{x^k\}$ converge vers un $x_\infty \geq 0$ avec $f(x_\infty) \leq f_\star + \epsilon$.*

Preuve :

(a) est évident.

(b) et (c) Nous allons premièrement vérifier les hypothèses du Théorème 6.2.1. Supposons que **Alg1** ne s'arrête pas. Alors, pour n'importe quel $x \in S_\epsilon$ nous avons

$$f(x^k) + \langle g^k, x - x^k \rangle \leq f(x) \leq f_\star + \epsilon = f^k_{\text{lev}} \quad \forall x \in S_\epsilon, \tag{6.8}$$

tel que pour $x \in S_\epsilon^\star$ nous avons $x \in H_k^- \cap \mathbb{R}_{++}^p$, $\forall k$, et donc (6.3)- (6.4) sont satisfaits. Donc, nous pouvons utiliser le Théorème 6.2.1, et la suite $\{x^k\}$ satisfait

$$x^k > 0, \quad \lim_{k\to\infty} \|x^k - E_{H_k}(x^k)\| = 0, \quad \lim_{k\to\infty} \|x^k - x\| = a(x), \quad \forall x \in S_\epsilon^\star.$$

Parce que la suite $\{x^k\}$ est bornée, il existe une sous-suite $\{x^{k_j}\}$ convergeant vers $x_\infty \geq 0$. De plus, par la définition de H_k nous avons

$$f(x^{k_j}) + \langle g^{k_j}, E_{H_{k_j}}(x^{k_j}) - x^{k_j}\rangle = f_\star + \epsilon. \tag{6.9}$$

Des hypothèses B, nous avons que $f(x^{k_j}) \to f(x_\infty)$ et g^{k_j} est bornée, donc passant à la limite dans la dernière égalité, nous obtenons

$$f(x_\infty) = f_\star + \epsilon \Longrightarrow x_\infty \in S_\epsilon.$$

De plus, de (6.8) $x_\infty \in H_k^- \cap \mathbb{R}_+^p$, $\forall k$. Et par conséquent,

$$a(x_\infty) = \lim_{k\to\infty} \|x^k - x_\infty\| = \lim_{j\to\infty} \|x^{k_j} - x_\infty\| \Longrightarrow a(x_\infty) = 0,$$

et donc l'ensemble de la suite $\{x^k\}$ converge vers x_∞. □

Maintenant, nous considérons le second algorithme **Alg2**, qui est obtenu à partir de **Alg**(f_{lev}^k) avec $f_{\text{lev}}^k := f(x^k) - \epsilon$, $\epsilon > 0$.

Théorème 6.2.3 *Soit $\epsilon > 0$ et supposons que l'ensemble optimal S est non vide. Alors $x^k \in H_k^-$ n'est jamais satisfait et nous avons que*

(a) **Alg2** *est bien défini.*

(b) Si **Alg2** *ne s'arrête pas, alors $\lim_{k\to\infty} f_{rec}^k \leq f_\star + \epsilon$, où $f_{rec}^k := \min_{j\leq k} f(x^j)$.*

Preuve :

La première affirmation est évidente parce que $f(x^k) \leq f_\star - \epsilon$ est impossible.

Pour prouver (a), supposons qu'à l'étape k l'algorithme ne s'arrête pas. Alors, nous n'avons jamais $g^k = 0$ ni $g^k \geq 0$ avec $\langle g^k, x^k\rangle \leq \epsilon$. Nous devons prouver l'existence d'un $x > 0$ avec $x \in H_k^-$ ce qui est équivalent à $\langle g^k, x\rangle \leq \langle g^k, x^k\rangle - \epsilon$.

- Si pour un indice i nous avons que $g_i^k < 0$, alors avec $x_i \to +\infty$, un tel x peut être trouvé.

- Si $g_i^k \geq 0$, $\forall i$, alors nous avons $\langle g^k,\, x^k \rangle > \epsilon$, et pour x suffisament petit, un tel x existe.

Par conséquent, **Alg2** est bien défini et (a) est prouvé.

(b) Preuve par l'absurde. Supposons le contraire ; alors $f(x^k) > f_\star + \epsilon$, $\forall k$ et pour chaque $x^\star \in S$, nous avons

$$f(x^k) + \langle g^k, x^\star - x^k \rangle \leq f_\star < f(x^k) - \epsilon,$$

tel que $x^\star \in (\cap_k H_k^-) \cap \mathbb{R}_+^p$ et donc toutes les hypotèses du Théorème 6.2.1 sont satisfaites. Dés lors, les suites $\{x^k\}$, $\{g^k\}$ sont bornées et $\lim_{k\to\infty} \|E_{H_k}(x^k) - x^k\| = 0$. Parce que

$$f(x^k) + \langle g^k,\, E_{H_k}(x^k) - x^k \rangle = f(x^k) - \epsilon, \tag{6.10}$$

par les hypothèses B, passant à la limite, il suit que $0 = -\epsilon$, ce qui impossible.
$\square$

Conclusion.

Le but de ce mémoire était de présenter les méthodes intérieures du type gradient et ϵ-sous-gradient pour la minimisation convexe avec contrainte.

Pour y arriver, nous avons premièrement établi un chapitre de préliminaire et un chapitre sur la méthode du point proximal.

Ensuite, nous avons un chapitre sur la méthode proximale logarithmique quadratique. Dans le chapitre quatre, nous développons la méthode de faisceaux intérieure, tout d'abord dans le cas général et ensuite dans le cas particulier où $C = \mathbb{R}^p_+$.

Enfin, les deux derniers chapitres sont consacrés à la description dans un premier temps des méthodes intérieures de descente du ϵ-sous-gradient avec quelques exemples et applications. Nous voyons aussi que les deux méthodes vues aux chapitres trois et quatre sont en fait des méthodes intérieures de descentes. Dans un second temps, nous avons une description de la méthode intérieure du sous-gradient où nous utilisons la projection logarithmique-quadratique.

Bibliographie

[1] Auslender A. et Teboulle M., *Interior gradient and epsilon-subgradient descent methods for constraint convex minimization,* Math programming 29, pp 1-26, 2004.

[2] Auslender A. et Teboulle M., *A log-quadratic projection method for convex feasibility problems.* D. Butanariu, Y. Censor, S. reich, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications,* Vol. 8. *Studies in Computational Mathematics.* Elsevier, Amsterdam, Pays-Bas, 2001.

[3] Auslender A. et Teboulle M., *Interior proximal and multiplier methods based on second order homogeneous kernels,* Math programming 24, pp 645-668, 1999.

[4] Auslender A., Cominetti R. et Haddou M., *Asymptotic analysis of penalty and barrier methods in convex and linear programming,* Math. Oper. Res. 22, pp 43-62, 1997.

[5] Bertsekas D.P., *Nonlinear Programming,* Athena Scientific, 1995.

[6] Hiriart-Urruty J.B. et Lemaréchal C., *Convex Analysis and Minimization Algorithms,* Springer, Berlin, 1996.

[7] Laurent P.J., *Approximation et Optimisation,* Hermann, Paris, 1972.

[8] Rockafellar R.T., *Convex Analysis,* Princeton, New Jersey, 1970.

[9] Strodiot J.J., *An introduction to nonsmooth optimization,* Cours donné à l'université des Sciences Naturelles, Ho Chi Minh Ville, 2005.